Theses and Dissertations | 1. Thesis and Dissertation Collection, all items

2017

Modeling environmental impacts on cognitive performance for artificially intelligent entities

Guthrie, Pierce C.

Monterey, California: Naval Postgraduate School

http://hdl.handle.net/10945/55610

NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

MODELING ENVIRONMENTAL IMPACTS ON COGNITIVE PERFORMANCE FOR ARTIFICIALLY INTELLIGENT ENTITIES

by

Pierce C. Guthrie

June 2017

Thesis Advisor:	Lee Sciarini
Co-Advisor:	Michael Guerrero
Co-Advisor:	Perry McDowell

This thesis was performed at the MOVES Institute.
Approved for public release. Distribution is unlimited.

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE		*Form Approved OMB No. 0704-0188*

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC 20503.

1. AGENCY USE ONLY (*Leave blank*)	2. REPORT DATE June 2017	3. REPORT TYPE AND DATES COVERED Master's thesis	
4. TITLE AND SUBTITLE MODELING ENVIRONMENTAL IMPACTS ON COGNITIVE PERFORMANCE FOR ARTIFICIALLY INTELLIGENT ENTITIES		5. FUNDING NUMBERS	
6. AUTHOR(S) Pierce C. Guthrie			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000		8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A		10. SPONSORING / MONITORING AGENCY REPORT NUMBER	

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government. IRB number ____N/A____.

12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release. Distribution is unlimited.	12b. DISTRIBUTION CODE

13. ABSTRACT (maximum 200 words)

The Marine Corps utilizes virtual simulations as a training tool for ground combat operations. Currently, the artificial intelligence of simulation entities does not exhibit appropriate performance degradation due to environmental conditions, such as heat and humidity. To address these gaps, this thesis reviews existing approaches to modeling the influence of environmental factors, specifically heat and humidity, on human performance in vigilance and attention tasks. We present a novel agent behavior model that incorporates a modified A* search pathfinding algorithm based on empirical evidence of human information processing under the specified environmental conditions. Next, an implementation of the agent behavior model is presented in a military-relevant virtual game environment. We then outline a quantitative approach to test the agent behavior model within the virtual environment. Results show that our human information processing–based agent behavior model demonstrates plausible agent performance degradation in hot, humid temperature environments when compared to paths around the danger area in mild temperature environments. The results of this research provide an approach for implementing an agent behavior model that accounts for environmental impacts on cognitive performance. We recommend future work to validate the model in a human subjects experiment to facilitate improving the realism of simulation training.

14. SUBJECT TERMS vigilance, attention, path finding, A* search, human information processing			15. NUMBER OF PAGES 137
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT UU

THIS PAGE INTENTIONALLY LEFT BLANK

MODELING ENVIRONMENTAL IMPACTS ON COGNITIVE PERFORMANCE FOR ARTIFICIALLY INTELLIGENT ENTITIES

Pierce C. Guthrie
Captain, United States Marine Corps
B.S., United States Naval Academy, 2011

Submitted in partial fulfillment of the
requirements for the degree of

**MASTER OF SCIENCE IN MODELING,
VIRTUAL ENVIRONMENTS, AND SIMULATION**

from the

**NAVAL POSTGRADUATE SCHOOL
June 2017**

Approved by: Lee Sciarini
Thesis Advisor

Michael Guerrero
Co-Advisor

Perry McDowell
Co-Advisor

Peter Denning
Chair, Department of Computer Science

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

The Marine Corps utilizes virtual simulations as a training tool for ground combat operations. Currently, the artificial intelligence of simulation entities does not exhibit appropriate performance degradation due to environmental conditions, such as heat and humidity. To address these gaps, this thesis reviews existing approaches to modeling the influence of environmental factors, specifically heat and humidity, on human performance in vigilance and attention tasks. We present a novel agent behavior model that incorporates a modified A* search pathfinding algorithm based on empirical evidence of human information processing under the specified environmental conditions. Next, an implementation of the agent behavior model is presented in a military-relevant virtual game environment. We then outline a quantitative approach to test the agent behavior model within the virtual environment. Results show that our human information processing–based agent behavior model demonstrates plausible agent performance degradation in hot, humid temperature environments when compared to paths around the danger area in mild temperature environments. The results of this research provide an approach for implementing an agent behavior model that accounts for environmental impacts on cognitive performance. We recommend future work to validate the model in a human subjects experiment to facilitate improving the realism of simulation training.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

ACT-R	Atomic Components of Thought-Rational
AI	Artificial Intelligence
CBRN	Chemical, Biological, Radioactive, Nuclear
CGF	Computer Generated Forces
CHAOS	Capability-based Human-performance Architecture for Operational Simulation
COMBATXXI	Combined Arms Analysis Tool for the 21st Century
CSR	Combat Stress Reactions
CRN	Common Random Numbers
C3HPM	Command, Control, and Communications Human Performance Model
DOD	Department of Defense
DSR	Distributed Soldier Representation
DVTE	Deployable Virtual Training Environment
ET	Effective Temperature
FPS	First Person Shooter
FTCST	Fire Team Cognitive Skills Trainer
HBR	Human Behavior Representation
HIP	Human Information Processing
IED	Improvised Explosive Device
IMPRINT	Improved Performance Research Integration Tool
IWARS	Infantry Warrior Simulation
MAGTF	Marine Corps Air Ground Task Force
MCRP	Marine Corps Reference Publication
MCSF	Marine Corps Synthetic Forces
MINDS	Modeling Individual Differences and Stressors
MOVES	Modeling, Virtual Environments, and Simulation
MTWS	MAGTF Tactical Warfare Simulation
ONR	Office of Naval Research
OneSAF	One Semi-Automated Forces

PMF	Performance Moderator Functions
PSF	Performance Shaping Functions
SAF	Semi-Automated Forces
SAMPLE	Situation Awareness Model for Pilot-in-the-Loop Evaluation
SCOPE	Soldier Capability Optimization for Projected Efficacy
SLATE	Soldiers Load Augmented Training Environment
USARIEM	U.S. Army Research Institute of Environmental Medicine
VBS	Virtual Battle Space
WBGT	Wet Bulb Globe Temperature

ACKNOWLEDGMENTS

First, I would like to thank my wife and best friend, Katherine, and my daughter, Lelia. I could not have completed this thesis, or my degree, without your love and support. Kat, thanks especially for your patience as you learned more about my thesis than you probably cared to—even with your own to work on! And thank you, Lelia, for always making me smile, and making every day wonderful for your mom and me.

I would like to thank my parents and family for their support through the years. I would not be here if they had not supported my dreams and put themselves second to their children.

I would like to thank my advisors for all of their hard work and patience in supporting me throughout the thesis process. Lee, thank you for weathering the long thesis meetings and reviews of this thesis. I have learned a ton from you and have grown so much from working with you during this process. I also enjoyed our Marine Corps bull sessions. Semper Fidelis. Perry, thank you for the conversation and the idea that sparked this thesis. Also, thank you for your insight, optimism, and for the extra push in motivation. Michael, thank you for taking time out of your busy schedule to add another task to the pile. You are an exceptional game designer, and I would never have coded this thesis without your help and mentorship.

I also would like to thank my cohort for their support and help throughout our time in MOVES. I learned so much from all of you beyond just our degree and will value each friendship always. Thanks also to Major John Wray for his help and patience in troubleshooting my model's design. Without your help and advice, this thesis would not have happened.

Finally, thank you, God, for all you have given me and for making this all happen.

THIS PAGE INTENTIONALLY LEFT BLANK

I.   INTRODUCTION

A.   BACKGROUND

The use of first-person shooter (FPS) simulations for training has become commonplace in the United States Marine Corps and United States Army. These training exercises can range from simple "terrain walks" to familiarize oneself with the operational environment to rehearsals via the simulation, and can provide valuable insight before execution (Lopez, 2014). However, in order to enhance the user experience and breed "buy in" to this training technique, the computer-generated forces (CGF) within the simulation must behave realistically.

An important characteristic of simulation is that it allows for situations in which trainees can train and practice individual, team and leadership skills without the need for other live participants. For example, training scenarios developed for the FPS Virtual Battle Space 2 (VBS 2) simulation allowed maneuver squad leaders to practice their decision-making skills in a study conducted in Camp Lejeune, North Carolina (Carroll et al., 2015). Traditionally, this type of training would occur in a live environment on a large training range. Furthermore, it would require participation from an entire squad, and consist of a relatively small number of trials due to time and logistics constraints. While simulation advances have helped address many of these issues, limitations in training system availability can limit simulation-based training events to a unit's key leadership. To this end, CGF powered by artificial intelligence (AI) may end up representing many of the subordinate members in a unit. Training with CGF powered by AI places an implicit necessity that these agents perform in a realistic manner that supports a high-fidelity training experience. The "goal" of CGF "is to develop these representations to the point that they will act automatically and with some realism without human intervention" (Pew & Mavor, 1998, p. 38). An observation made in an Office of Naval Research (ONR) presentation on an AI virtual simulation training suite highlighted the need for realism, declaring that "unacceptable role player AI within existing simulations and game environments prevent widespread adoption" of those simulations for training use (Stensrud & Hamel, 2015, slide 2).

Many of the operating environments that the Marine Corps currently operates in have high levels of heat and humidity. This environmental reality can adversely affect the physical and cognitive performance of Marines who are required to operate in such conditions. Partially addressing these realities, important advances in the physical CGF models within FPS simulations have been made in recent years that account for certain physical performance decrements. Specifically, the U.S. Army's Virtual Battle Space 3 (VBS 3) simulation has made significant strides in the realism of the physical models utilized for FPS simulations (Lopez, 2014). VBS 3 allows Soldier avatars to model the physique and physical performance of the actual Soldier by accounting for the Soldier's "height, weight, Army Physical Fitness Test scores and even their weapons qualifications scores" (Lopez, 2014, paragraph 2). These inputs create more realistic-looking avatars, and more importantly, generate decremented marksmanship performance and fatigue of the Soldier's avatar based on the avatar's physical actions in the simulation (Curthoys, 2014). Furthermore, VBS 3 allows exercise designers to "manipulate different weather patterns" ("Army Gaming releases newest training product," 2014, p. 2). The physical models within military simulations offer cutting-edge visualization of the military operating environment, but a gap exists in modeling the effects of the physical environment on the cognitive skills internal to each Marine or Soldier represented within the simulation.

1.   Research Questions

The realism of a FPS military simulation can greatly improve if it can model the impacts of the environment on certain cognitive aspects of the CGF entities. This thesis addresses several key research questions with respect to improving the modeling of the environment within training simulations. First, to what extent are the environmental factors of heat and humidity accurately modeled in FPS and constructive military simulations? And if modeled correctly, do these environmental factors contribute to realistic performance of the AI, or behavior, of the simulated entities within the simulation? Next, to what extent can an investigation of human performance of tasks relevant to military operations, such as vigilance and attention, executed under various

environmental conditions be used to quantify and improve models used to create the AI of entities executing a similar task in a military training simulation?

This thesis examines legacy and current military simulations with respect to the extent to which they model environmental impacts on cognitive processes. We will explore both FPS and constructive types of military simulations to gain a better understanding of ways in which environmental modeling techniques could improve the behavior in both types of simulations. For example, it is suggested that constructive simulations could use more realistic behavioral input (Blais, 2016). Wing (2002) pointed out that "we need to devise a means to inject human-behavior-representation-in-the-loop in constructive simulations" (p. 27). Aligned with this call to action, this thesis will first investigate current approaches to modeling pathfinding in CGF. Next, this thesis will analyze two key aspects of cognition, attention and vigilance, within a human information-processing model proposed by Wickens, Hollands, Banbury, and Parasuraman (2013). Attention and vigilance will be discussed within the broader context of multiple studies that analyzed the influence of heat and humidity on these psychological constructs. Finally, this thesis proposes a model of a CGF agent executing pathfinding in an improvised explosive device (IED) danger area environment. This model will demonstrate how simulation designers could modify a CGF agent's AI to reflect cognitive decrement in vigilance and attention related tasks as a result of environmental heat and humidity.

2.   Performance Capabilities Needed for CGF

CGF in military simulations must possess many qualities such as physical appearance and mental processes taking place within the CGF's "mind." The efforts of the *Panel on Modeling Human Behavior and Decision Making: Representations for Military Simulations* provide an extensive description of human behavior modeling, with specific focus on several cognitive properties and constructs that the panel felt would lead to increased realism of military simulation (Pew & Mavor, 1998). It should be noted that the authors mention the military user audience as a key constituent that must be able to trust the validity of the models they use. Furthermore, their report provides an argument

for grounding human performance models in the results of experimental data, and for how these data can range from operational performance data to inherent human behavioral characteristics. Pew and Mavor's (1998) contribution offers a broad span of potentially important cognitive factors and constructs for inclusion within human performance models, such as situational awareness[1] and individual decision-making within the realm of command and control. Although these topics fall outside of the scope of this thesis, they are worth mentioning given the fact that the authors make several recommendations for modeling these human qualities via decremented functioning of attention or human information processing. We will similarly influence attention and human information processing within this thesis.

Ultimately, this thesis aims to offer an approach that will enable increased fidelity of military simulations, as well as work towards achieving the goals of Schroth (1989) who emphasized the need to include representation of human factors in combat models. In his work, Schroth suggests sensitivity analyses be used to assess the influence that human factors can have on combat models, and provides a framework to use artificial intelligence for the purposes of achieving this goal.

a.   Attention

In their work, Pew and Mavor (1998) included tactical vignettes of military personnel in mission scenarios in order to provide a tangible way of grasping the detail inherently needed in simulation human behavior models. At the time of their publication, the authors identified a general dissatisfaction with the realism of human behavior offered in military simulation. More importantly, the authors emphasize that cognitive properties and constructs exist, attention for example, that can help create more realistic human models. Furthermore, the authors suggest that training system developers should adopt a core notion that "movement from the current state of human behavior representation to the achievement of higher levels of realism with respect to observable outcomes requires significant understanding and application of psychological and organizational science"

---

[1] For a detailed description and discussion of situational awareness, see Endsley (1995).

(p. 14). The inclusion of attention as a valuable CGF property to model aligns with the quest to improve CGF human behaviors, and serves as validation for studying how to better model attention within CGF entities. Notably, the authors point out that "divided attention and multitasking—doing several things at once—are ubiquitous in combat operations" (p. 112).

b.   *Behavior Moderators*

Mental and physical "behavior moderators," such as environmental heat, comprise one of many "external moderators" of human performance recognized as important factors for realistic CGF (Pew & Mavor, 1998, p. 245). In their work, Pew and Mavor (1998) attribute Van Nostrand (1986) with leading efforts to review and encompass numerous moderators of human performance into combat models. The authors also commend Van Nostrand for seeing the value of empirical data for modeling Soldiers operating on the battlefield in stressful circumstances. Van Nostrand's (1986) work is foundational because it calls for the inclusion of actual empirical data derived purposefully for use in combat models instead of repurposing data from other uses or studies. This idea finds traction in a scenario envisioned by NATO where military units in the year 2020 will train with repositories of human behavior data that provide the ability to combine and utilize "models such as thermal physiological response to working in hot environments coupled with vigilance and decision making models that are moderated by body temperature and hydration status" (North Atlantic Treaty Organization (NATO), 2009, p. 1–8). Notably, Pew and Mavor (1998) point out that data collection of varying human performance proves anything but easy as a result of the various conditions exhibited by operational environments and missions.

Another key takeaway from Van Nostrand (1986) and reinforced by Pew and Mavor (1998) includes the observation that multiple characteristics of human combat performance fall under the influence of a multitude of moderators (including heat and humidity); thus, one must realize that no combat model will prove complete, but value can still come from having human performance modeled and negatively influenced in some way. Van Nostrand uses a specific example of how to model heat's effect on a tank

crew, describing possibilities of accounting for individual variance in performance, as well as changing the rate of declining performance for the crew members.

Pew and Mavor (1998) make note of a quote from Van Nostrand's (1986) work commenting on previous Chemical, Biological, Radiological, and Nuclear (CBRN) performance studies. Van Nostrand's observation proves at least tangentially related to heat's impact on the performance of military personnel:

> Apparently the major problems are the heat buildup inside the clothing (in which case performance decrements associated with high temperatures should be expected), the inability to perform work requiring manual dexterity with the heavy gloves on, and to see well with the face mask in position. The end result of prolonged exposure to heat is extreme fatigue, and the result of fatigue is degradation of thinking and decisionmaking (*sic*) skills. Therefore, the tests that require soldiers to perform tasks that are so well practiced that they can do them without thinking about them will show less decrement than tasks which require the soldiers to decide what must be done next. Any task which requires vigilance will probably show large performance decrements. (Van Nostrand, 1986, pp. 2–10)

Fatigue as an aspect of declining human performance in combat falls outside of the scope of this thesis, but Van Nostrand's connection of heat exposure to decision-making issues and vigilance is an important concept for our proposed model.

c.   *Human Data in Simulations*

The U.S. Army has conducted extensive research into the performance capabilities and characteristics required of human performance models within military simulations. An example of this research entails an effort to build a human performance repository known as the Distributed Soldier Representation (DSR) designed to serve as a mechanism for implementing realistic performance data within simulated CGF (Fefferman et al., 2015). One of the categories of performance data specifically outlined within the DSR report centers around the effects of stress on performance (Fefferman et al., 2015). In 2012, the U.S. Army Health Promotion and Wellness website article titled

"A Soldier's Guide to Deployment Related Stress Problems," linked stress from combat[2] as manifesting itself in varying forms to include physically quantifiable results such as "fatigue, slower reaction times, indecision, disconnection from one's surroundings, and inability to prioritize" (as cited in Fefferman et al., 2015, p. 14).

Fefferman et al. (2015) utilize Marine Corps Reference Publication (MCRP) 6–11C *Combat Stress* (2000) to link heat stress to training and combat. MCRP 6–11C (2000) points out that combat can result in "exposure to heat, cold or wetness" for ground combatants (p. 2). The MCRP also points out that "environmental stressors often play an important part in causing…adverse or disruptive combat stress reaction behaviors" (p. 2). Fefferman et al. (2015) mention the Department of Defense (DOD) simulations One Semi-Automated Forces (OneSAF), Infantry Warrior Simulation (IWARS), and VBS 2 as simulations where one could model combat stress within entities. The authors also offer several ways to mimic the effects of combat stress within a simulated Soldier entity. A few proposed examples for simulating combat stress within Soldier entities derived from a guide titled "Leaders Guide for Managing Marines in Distress" include longer (i.e., slower) reaction times, and an "inability to perceive (cannot see/hear)" (as cited in Fefferman et al., 2015, p. 98). The DSR report implies a need for modeling of environmental factors such as heat in yielding combat stress reactions (along with many other sources of stress), and specifically identifies a need for modeling indicators of combat stress reactions (CSR) within military simulation entities (Fefferman et al., 2015).

Fefferman et al. (2015) point to the work of Steadman (2011) to link stress and attention. Steadman suggests that "stress degrades the form of conscious attention known as 'working memory'" (Steadman, 2011, Abstract). Fefferman et al. (2015) created an implementation of their ideas by combining an application known as Effects of Stress (EoS) as a basis for a DSR server that would allow a simulation to "degrade the respective Soldier's performance" in something "such as…small-arms fire accuracy"

---

[2] The DSR report covers stress under the more general heading of Combat Stress Reactions (CSR) described in a U.S. Army Medical Department report called *Combat Stress* as "expected and predictable emotional, intellectual, physical and behavioral reaction from exposure to stressful events" (as cited in Fefferman et al., 2015, p. 90).

based upon the Soldier's "unique, dynamic overall stress level" (p. viii). Ultimately, Fefferman et al. (2015) identify a modeling gap for having "CSR symptoms…instantiated within individual entities" (p. 97).

Fefferman et al. (2015) also point to the work of Steadman (2011) for describing why physiology receives focus in their report. Steadman (2011) describes how "physiological response to combat can degrade…cognitive capability" for leaders "during stressful situations" (Steadman, 2011, Abstract). Furthermore, Fefferman et al. (2015) classify environmental heat and humidity primarily in the physiological domain. The authors make a case for including physiology in simulations to improve realism, identifying a common simulation weakness that "simulated Soldiers in current large-scale simulations never become fatigued" (p. 51). These researchers also combined the DSR server and the Army Research Lab Soldier Load Augmented Training Environment (SLATE)—a teaching tool for Soldiers to learn how to properly load out their carried equipment. The broader project that SLATE aggregates to, the Dismounted Solder Centric Load and Route Planning Mobile Training Applications, incorporates physiological heat strain and "energy expenditure models," as well as "route analysis" (Fefferman et al., 2015, p. 50).

Santee, Reardon, and Pandolf (2012) provide an extensive description of some of the physiologically focused biomedical models developed by the U.S. Army Research Institute of Environmental Medicine (USARIEM), and how laboratory study generated empirical data plays a crucial role in USARIEM modeling research. The authors also describe the physiological nature of inputs and outputs of the USARIEM heat strain model.[3] Pandolf, Stroschein, Drolet, Gonzalez, and Sawka (1985) have created a version of the USARIEM model that considers "ambient air temperature… and relative humidity" (p. 13) as two possible inputs, along with "deep body…temperature and sweat loss" (p. 2). This model produces "predicted" outputs that include the "expected physical work-rest cycle, the maximum single physical work time…and the associated water

---

[3] Santee et al. (2012) reported that the heat strain model has taken many forms, and they provide a history and overview of the model.

requirements" (p. 2). The authors provide a description of the underlying models and equations within their implementation of the USARIEM heat strain model. One of these underlying equations created by Pandolf, Givoni, and Goldman (1977) focuses on standing and walking, and resulted from a study that found how "energy expenditure increased with external load, both standing and walking" (Abstract). Ultimately, Santee et al. (2012) provide additional support to the DSR effort by arguing that "physiologically sound models can also be used in simulations that help to develop video games used as training tools" for military leadership (p. 40).

In sum, this thesis aims to model CGF behavior modifications by accounting for environmental heat stress' effects on vigilance and attention. Ultimately, the resultant effects on entity decision-making in choosing a route around potential IED danger areas will be demonstrated by linking human performance data with human behavior model performance to cause performance changes as proposed by Fefferman et al. (2015).

B.   CURRENT STATE OF CGF AND INTELLIGENT AGENTS

Simulation for training and analysis encompasses a broad range of tools and techniques for implementing human behaviors within CGF. An analysis of current and former constructive and FPS[4] simulations and their implementation methods of CGF human behaviors in the presence of environmental factors follows.

1.   Virtual Battle Space Human Behavior Modeling Approaches

Both VBS versions 2 and 3 exist as a part of the United States Marine Corps' Deployable Virtual Training Environment (DVTE)—a set of systems that incorporates multiple training simulations for the purpose of training and sustaining "individual, team, and unit critical war fighting cognitive skills" for "combined arms, squad, and platoon level tactics" (Program Manager Training Systems, 2016, p. 19). An early version of the DVTE's use of VBS (called the *Virtual Battlefield Simulation* at the time) was in the Fire Team Cognitive Skills Trainer (FTCST) that promoted CGF manipulation by the Marine

---

[4] Pew and Mavor (1998) call these "virtual simulations" (p. 34).

trainees (Bailey, 2002). An important characteristic offered by VBS's environment is the capability for an instructor/operator to modify the weather and how constructive CGF behave (Program Manager Training Systems, 2016). The following section provides an assessment of the capability and performance of friendly CGF in VBS 3.

a. Virtual Battle Space 3 Experimentation

A summary investigation into VBS 3's modeling and CGF AI capabilities showed it is not possible to manipulate the temperature of the game environment, although, weather effects such as fog, rain, and snow can be implemented (Bohemia Interactive Simulations, 2015). The VBS 3 manual reports the AI of the agents within VBS 3 can be influenced as far as behavior in terms of combat posture, ranging from "careless" to a more stringent "combat" mode of behavior that changes while accounting for tactical considerations such as cover locations, or alters behavior based upon enemy presence. Further, the agent's pathfinding has the ability to account for enemy threats (Bohemia Interactive Simulations, 2015).

VBS 3 has extensive modeling capabilities that influence the CGF's fatigue and morale, and how those factors influence the CGF's performance—for example, the VBS 3 manual reports that morale and fatigue serve to cause "'tunnel vision' and decreased accuracy" (Bohemia Interactive Simulations, 2015). Currently, VBS 3 implements adjusted versions of Pandolf running and walking equations for fatigue modeling (Wojtowicz, 2013).[5] The manual also reports that one can also edit a CGF unit's level of training and endurance—both of which also effect the modeling of CGF fatigue and morale. VBS 3 also allows scenario editors to modify a unit's navigation capability, and adjust it to induce errors that impact the unit's arrival at the correct location of a waypoint (Bohemia Interactive Simulations, 2015).

---

[5] See Wojtowicz (2013) for their explanation of these equations.

b.   *Virtual Battle Space 2 and CoJACK Human Behavior Construct*

Evertsz, Pedrotti, Busetta, Acar, and Ritter (2009) created a simulation scenario within VBS 2 that included a cognitive architecture called CoJACK to drive the AI of a suicide bomber and civilian populace agents. Evertsz, Ritter, Busetta, and Pedrotti (2008) describe CoJACK as a type of cognitive architecture known as a Belief, Desire, Intention (BDI) architecture, and cognitive architectures as generally "a computational system that models the structural properties of the human cognitive system—the information processing mechanisms that are fixed across tasks" (p. 2). CoJACK was designed for the purposes of modeling variability between individuals as a result of their cognitive, perceptual, physiological, and motor skill differences combined with the effects of behavior moderators resulting from external, internal, and task-caused conditions (Ritter & Norling, 2006).

The linking of CoJACK into the AI of VBS 2 came from the need to upgrade the realism of CGF behavior beyond just path finding and script execution (Evertsz et al., 2009). The authors note some of the useful aspects of pathfinding consist of its ability to communicate to users a CGF's behavioral variability, as well as a means for allowing "intelligence to play out…and an area for affect to influence" (Evertsz et al., 2009, p. 7). The authors also offer tips on how to best mesh and de-conflict pathfinding issues that arose between VBS 2's AI functionality and CoJACK's implementation.

Evertsz et al. (2009) aimed to delineate VBS 2 and CoJACK based on a separation where CoJACK would handle the internal aspects of an agent—including physiology—while VBS 2 would serve as the "mediator of perception, action, and environmental influences such as terrain and temperature" (p. 1). Unfortunately, the authors did not report the extent to which VBS 2 effectively models temperature. Evertsz et al. (2009) achieved integration of the cognitive architecture and simulation by porting CoJACK into VBS 2 via the simulation's Application Scripting Interface (ASI)—essentially a translation mechanism for outside AI to communicate and modify the behaviors that take place within VBS 2.

Evertsz et al. (2009) do discuss endurance and fatigue as possible VBS 2 CGF parameters that could be modified by external AI. The focus in their iteration of CoJACK and VBS 2 integration relied on having fear and morale serve as the main behavior moderators. These authors use the ASI and inherent VBS 2 parameters as a mechanism for altering the performance of the CGF within VBS 2. Further, the CoJACK architecture supplies an additional source of variability by providing unique cognitive performance abilities for each CGF, i.e., not all CoJACK-sponsored CGF will have the same ability to process information (Evertsz et al., 2008; Evertsz et al., 2009). Evertsz et al. (2009) describe a number of AI and scripting issues they had to overcome in order to de-conflict AI architectures and built-in functionality between CoJACK and VBS 2 capable of achieving the desired behaviors. Ultimately, this implementation example showcases an advanced mechanism for CGF AI.

2. **Discrete Event and Constructive Simulation Human Behavior Modeling Approaches**

The DOD utilizes other simulation types beyond FPS-style simulations for achieving training or analysis objectives, including discrete event simulations and constructive simulations. The following simulations are relevant to a discussion of the implementation of human behavior models in CGF. We also pay specific attention to how they model the environment and its relational impact on CGF cognitive performance.

a. *Improved Performance Research Integration Tool (IMPRINT)*

The U.S. Army's Improved Performance Research Integration Tool (IMPRINT) constitutes an important model for discussing human performance models that have mechanisms for implementing performance decrement due to various environmental characteristics (United States Army Research Laboratory, n.d.). IMPRINT is a discrete event simulation, and provides a key usage capability for determining how well system operators can interact with a system even while performing under the influence of "environmental stressors" (United States Army Research Laboratory, n.d., paragraph 2). The mechanism for modeling these environmental stressors entail "embedded algorithms" (United States Army Research Laboratory, n.d., paragraph 6). The structure

of IMPRINT consists of a "human-system task network modeling tool" (Allender, 2000, p. 140).

Allender (2000) describes the IMPRINT model as a human performance model used primarily in the U.S. Army's acquisitions process to promote the best possible design of a system. The author also describes IMPRINT as a means to improve a system's performance and serve as a cost-mitigation technique within the overall acquisitions process (i.e., in the long run it proves cheaper to model humans via a computer than live in the field). As a relevant example of IMPRINT's uses, Allender (2000) describes work accomplished by Perry, Davis, and Fields where IMPRINT supplied inputs to a constructive analysis model called the Combined Analysis and Support Task Force Evaluation Model (CASTFOREM) to "help predict high-level force-on-force effectiveness" (as cited in Allender, 2000, p. 143).

We refer the reader to the Chapter VI supplemental to gain a better understanding of the functionality and execution of the model within IMPRINT, as it inspires certain features within our model.

b. Agent Cognitive Architectures and Performance Shaping Functions

Numerous sophisticated approaches exist to model varying human performance within CGF that span a range of complexity. For example, Gillis and Hursh (1999) offer a complex implementation of a model that varies human performance as a result of combining factors of generic battlefield stress and fatigue, along with more interesting factors such as intelligence, level of experience, and type of personality for modeling a commander's decision-making in a simulated National Training Center battle based on real data. We now provide a brief overview of more sophisticated approaches, such as cognitive architectures, cognitive-affective architectures and performance shaping functions, to modeling human behavior in the realm of human information processing (Zacharias, MacMillan, & Van Hemel, 2008).[6]

---

[6] Zacharias, MacMillan, and Van Hemel (2008) offer a thorough summary of these types of modeling approaches.

(1)   Cognitive Architecture Modeling Possibilities

Ritter and Avraamides (2000) provide a relevant foundation for this thesis with respect to thinking about how to implement human performance moderators within models. The authors, referencing Boff and Lincoln, point out that "temperature" and "humidity…all influence attention and problem solving" (as cited in Ritter and Avraamides, 2000, p. 4). Building on this observation, the authors point out a weakness in simulated virtual environments in that they model "change in the physical environment," but do not model how these changes affect agents operating within the environment (p. 4).

Ritter and Avraamides (2000) classify behavior moderators as extrinsic, intrinsic, and task based, and place them within a stage model of human information processing that appears quite similar to the one put forth by Wickens et al. (2013). The authors propose meshing their architectural model construct with a simulation such as ModSAF via an external architecture such as Soar, Atomic Components of Thought-Rational (ACT-R), or Jack (a precursor to CoJACK). An important aspect of the architectural model proposed by Ritter and Avraamides (2000) would include multi-tasking, since multi-tasking ensures that "attention and task switching" get "exercised" (p. 21).

An important contribution of the work by Ritter and Avraamides (2000) comes from their enumeration of the many behavior moderators required, and how these moderators could impact agent performance. The authors even provide use cases with operational examples. For example, the authors state that "temperature and exposure time" will have an effect on vigilance and hinder the ability to make faster decisions, resulting in an implementation where "a team that is exposed to excessive heat for a long period of time will not respond as fast to sudden threat such as an ambush" (p. 11). The authors also link the moderator of humidity to fatigue, and describe how the accumulated effects of exposure to elevated temperatures can cause fatigue. Furthermore, Ritter and Avraamides (2000) detail how agents experiencing fatigue could "choose low effort/low probability of success strategies over high effort/high probability options" (p. 19). This thesis has chosen to focus on heat's relationship to an agent's performance in vigilance and attention-type tasks, but Ritter and Avraamides' (2000) linking of these

environmental characteristics to fatigue showcases the multitude of human performance characteristics that one could model.

Ritter and Avraamides (2000) also include a section discussing how to manipulate perception (classified as internal moderators) within agents. They offer an example where an agent's "perceptual accuracy" being affected could cause the agent to "spend more time…on searching for targets" (p. 12). Furthermore, an agent's perception of threats to itself could be influenced whereby they will direct their attention more to enemy-detecting sensors that indicate a pending attack (Ritter & Avraamides, 2000). The authors even describe how novices can be depicted in simulation by displaying them as possessing "slower processing speed, slower movement speed" and in making "more errors" (p. 8).

Applying Ritter and Avraamides' (2000) own logic of the interactive nature of behavior moderators, the properties of perception could also be negatively influenced by a number of factors, including temperature, resulting in agents not necessarily paying appropriate attention to searching for targets or recognizing IED danger areas. While, the recognition that simulation realism could increase by modeling stress increasing as a result of modeling "weather conditions" at the individual agent-level has existed in the literature (Ritter et al., 2003, p. 56), the capability has been virtually nonexistent. Another classification of internal moderator, called "action," could show effects with respect to the accuracy or inaccuracy of an agent's movement as a result of interacting with attention and other moderators relating to "task, anxiety" and "fear" (Ritter & Avraamides, 2000, p. 14). The authors include figures that visually depict the relationship between moderators and an agent's performance (see Figure 1 and Figure 2).
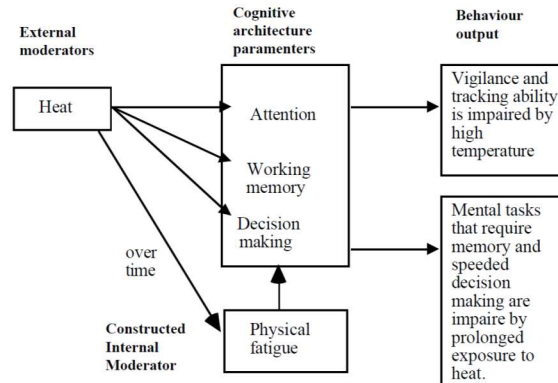
Figure 1.   Model of Temperature and Exposure Time Moderators.
Source: Ritter and Avraamides (2000).
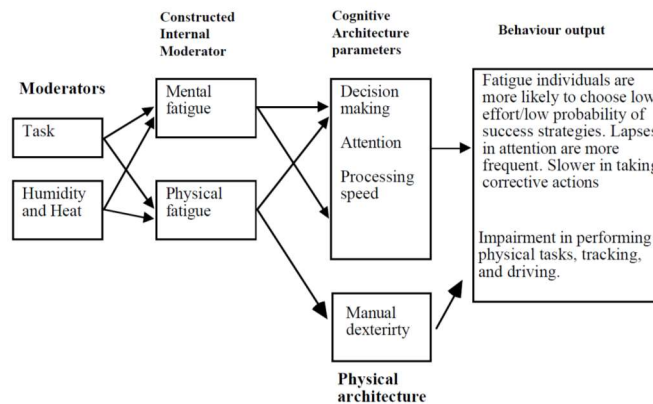
Figure 2.   Model of Humidity, Heat, and Fatigue Moderators.
Source: Ritter and Avraamides (2000).

(2)   Performance Shaping Functions (PSFs)

Paulsen, Alicia, and Shrader (2012) describe an effort to quantify environmental stressors via algorithms known as performance shaping functions (PSFs) for integration into modeling systems such as IMPRINT. An important aspect of this project focuses on the methodology to incorporate data from the literature and experimental efforts into the creation of the PSFs, and then to validate the outputs of the PSF-supported models against "human in the loop" data acquired via experimentation (p. 976). The project focused on motion as a stressor within a broader model of U.S. Navy shipboard tasks, but did focus heavily on "thermal stress" during their data acquisition phase (p. 977). In

general, the motivations of the authors centered on providing a more general ability to incorporate multiple environmental stressors that possess adequate "predictiveness" into models (p. 979).

We also reviewed the PMFserv architecture, which potentially serves as a means to counter trends observed by the authors where agents possess super-AI "capabilities that no real human being would possess," unencumbered by the "effects of fatigue, stress, heat…or other factors that would likely affect the performance of a real human operator" (Silverman, Johns, Cornwell, & O'Brien, 2006, p. 3). PMFserv is described as a human behavior architecture consisting of performance moderator functions (PMFs) that are essentially different "theories and models" describing human behavior (Silverman et al., 2006, Abstract). PMFserv consists of multiple modules within an agent's representation of itself including modules for perception and "biology/stress," along with other modules to represent other elements of human behavior such as a "personality, culture, emotion" module (Silverman et al., 2006, p. 7). An interesting quality of Silverman et al.'s (2006) work entails the inclusion of other models as PMFs within PMFserv, such as their own version of the Gillis-Hursh (1999) model and inclusion within the bounds of that model of "reservoir tanks" to represent sources of fatigue (p. 10). The authors give a list of the numerous efforts to utilize PMFserv mostly in simulations of crowd behavior, including their own implementation to simulate "crowd and militia behaviors observed in the Ranger operation in Mogadishu" by integrating PMFserv with Unreal Tournament (Silverman, Johns, et al., 2006, p. 26).[7]

(3)　　Cognitive Architecture Modeling- SOAR and ACT-R

Soar (Laird, Newell, & Rosenbloom, 1987) and ACT-R (Anderson & Lebiere, 1998) comprise two other types of cognitive architectures that have found use in driving human behavior models within CGF. Pew and Mavor (1998), referencing Newell, describe Soar as a "symbolic cognitive architecture that implements goal-oriented behavior as a search through a problem space and learns the results of its problem

---

[7] See Silverman, Bharathy, O'Brien, and Cornwell (2006) to learn about this PMFServ integration with Unreal Tournament in more detail.

solving" (as cited in Pew & Mavor, 1998, p. 90). Since being introduced, Soar has found practical application in the real world, and in human behavior cognition modeling (Pew & Mavor, 1998). Kelley, Patton, and Allender (2001) describes ACT-R as a "symbolic, production system architecture, capable of low-level representations of memory structures" (p. 1455). A project describing the linkage of ACT-R and Unreal Tournament describes ACT-R as a "high-fidelity architecture" capable of providing models of the simplest of human cognitive behaviors to more involved tasks such as "high-level decision-making" (Best, Lebiere, & Scarpinatto, 2002, p. 34).

Multiple use cases of ACT-R and Soar demonstrate its relevance to modeling the environment within simulation. One such use case connected a version of ACT-R, known as ACT-RΦ, with a simulation called HumMod (Dancy, 2013; Dancy, Ritter, & Berry, 2012). HumMod, the work of Hester et al., can be described as a human physiology simulation that can model heat within humans (as cited in Dancy, 2013). Dancy, Ritter, and Berry's (2012) goal for this capability entailed providing a means for linking "environmental…effects on central and peripheral physiology" to show how these would influence parameters representing cognition (p. 81).

Allender, Kelley, Lockett, and Archer (2005) describe the applicability and flexibility of the IMPRINT model in integrating with other types of models and simulations, such as ACT-R. The authors describe the significance of this integration as allowing for much of the human performance data that drives IMPRINT to come from the ACT-R model rather than from consultations with subject matter experts, literature reviews, or field tests. Allender et al. (2005) describe how the U.S. Army integrated ACT-R with its modeling efforts of Soldier situational awareness—describing the work of Kelley, Patton, and Allender (2001), who "used it to predict the errors in situation awareness of infantry soldiers while navigating cross-country on foot" (p. 1194).

Allender et al. (2005) discuss Pew and Mavor's (1998) work as a driver of integrating models with cognitive models such as ACT-R. The authors envisioned models such as IMPRINT inheriting a mechanism for modeling "attention, memory, and detailed cognitive processing" (Allender et al., 2005, p. 1193). Allender et al.'s (2005) use of a few examples relating to situational awareness leads us to believe that they likely refer to

attention as pertaining to situational awareness as described by Endsley (1995), however, we would like to note that attention does not rely on situational awareness to exist as a cognitive property.

Wray, Laird, Nuxoll, and Jones (2002) outline their plans and previous work for applications of Soar to the FPS game platform Unreal Tournament as a part of an Office of Naval Research (ONR) program. Their work provides an intriguing implementation of environmental influence on CGF. An example of Soar integration with Unreal Tournament in a game called *Haunt 2*, links environmental temperature to fatigue, exertion, and body temperature level of the CGF (Laird et al., 2002). The authors describe how a decline in temperature has a direct influence in causing the CGF to become more fatigued. Wray, et al. (2002) make note that they remain undecided on which environmental influences seen in *Haunt 2* that they may use within the ONR project's CGF, but that the mechanisms for doing so remain in place due to the value added in terms of more realistic and variable behavior.

(4)   Other Cognitive Architectures

Ubink, Aldershoff, Lotens, and Woering (2008) describe work accomplished in the Netherlands called the Capability-based Human-performance Architecture for Operational Simulation (CHAOS) that serves as the underlying cognitive architecture for the Soldier Capability Optimization for Projected Efficacy (SCOPE) infantry analysis model. The CHAOS architecture follows a model of pandemonium theory—essentially different human behaviors engage in a competition for human performance characteristics (called resources)—where only a certain amount of these resources are available (Ubink et al., 2008). Attention serves as one of the human performance resources that the behaviors could compete over, and stress plays a role in potentially impacting these resources (Ubink et al., 2008).

Ubink et al. (2008) also describe an implementation of CHAOS where the Dutch and Afghanistan weather climates influenced the performance of agents within SCOPE executing a peacekeeping operation. The authors describe how the agents took longer to finish the operation in the hotter climate. SCOPE and CHAOS incorporate several

models to account for the carried load of Soldiers and acclimatization's role in varying performance in the heat (Ubink et al., 2008; Ubink, Lotens, & Aldershoff, 2008). The model also implements fairly detailed physiology models that account for environmental climate metrics, such as temperature and humidity, although admittedly SCOPE remains limited because of its inability to model visual perception, or to link vision to the resources up for competition (Ubink et al., 2008; Ubink, Lotens, & Aldershoff, 2008). SCOPE also models threat perception as a part of a situational awareness model (Ubink et al., 2008). The CHAOS implementation described by Ubink et al. (2008) illustrates this threat perception behavior. Ultimately, the contribution of CHAOS is an example of a general framework capable of allowing environmental stressors to influence cognitive performance in a human characteristic such as attention and overall threat perception (Ubink et al., 2008).

c.   *Semi-Automated Forces*

Constructive simulations often make use of semi-automated forces (SAF), described by Pew and Mavor (1998) as "virtual simulations of multiple objects that are under the supervisory control of a single operator" (p. 40). A few key military simulations employing SAF include ModSAF and OneSAF. Pew and Mavor (1998) describe a ModSAF derivative in a project from the 1990s known as Marine Corps Synthetic Forces (MCSF). The authors, referencing Hoff, describe MCSF as a simulation that delivered CGF models for Marines at the squad level and below for the "purposes of training their respective superiors" (as cited in Pew & Mavor, 1998, p. 41). SAF decisions within MCSF are made based upon choosing from tables of ranked decisions for a given situation and simulation state (Pew & Mavor, 1998). However, MCSF did not try to implement higher-level cognitive abilities within CGF agents, including "information gathering, perception, correlation, or situation assessment" (p. 41).

Pew and Mavor (1998) describe ModSAF as fairly primitive with respect to human behavior modeling. Downes-Martin enumerate only a few basic behaviors for ModSAF such as "move, shoot, sense, communicate, tactics, and situation awareness" (as cited in Pew & Mavor, 1998, p. 40). In fact, Pew and Mavor (1998) argue, "It is

impractical to use ModSAF to construct general-purpose behavioral or learning models," because "no underlying model of human behavior" exists (p. 40). This results in "any behavior representation" needing to "be coded into the finite-state machine" of ModSAF (p. 40). Furthermore, the authors relate that a burdensome human-in-the-loop requirement exists for controlling ModSAF agents. One of the main reasons for SAF control and intervention by humans stems from a need to correct realism deficiencies in the CGF for the purposes of better training (Abdellaoui, Taylor, & Parkinson, 2009). Reece, Kraus, and Dumanoir (2000) describe more advanced path finding implementations within ModSAF in order to create tactical path finding and movement of infantry CGF agents. Interestingly, these authors chose a movement technique called "cell decomposition;" this technique is akin to grid search representations that use A* path finding algorithms (Reece, Kraus, & Dumanoir, 2000, p. 5). Another more advanced implementation of ModSAF includes instances of using Micro Saint models (a tool utilized within IMPRINT) to generate human performance data describing declining performance in terms of both timing and accuracy of task completion that were then fed into ModSAF simulations (Pew & Mavor, 1998).

OneSAF is an entity-level simulation capable of providing CGF within constructive simulations or virtual training simulations utilized by the U.S. Army and U.S. Marine Corps (G-3 Operations and Training, n.d.; PEOStri, n.d.). OneSAF employs a weather modeling capability that can model temperature and relative humidity (Pfeiffer & Tamash, 2014). These authors describe an experiment with OneSAF that demonstrated significant improvements to simulation realism with respect to ground convoy and air operations of a heavy March 2003 dust storm during the Second Gulf War in Iraq. The improved realism was manifested in longer times to complete missions (Pfeiffer & Tamash, 2014). The U.S. Army also led a research effort that resulted in ACT-R's implementation with the OneSAF Testbed Baseline[8] for the purposes of modeling a commander's situational awareness (Allender et al., 2005; Juarez-Espinosa & Gonzales,

---

[8] OneSAF Testbed is a development environment for a precursor version of OneSAF called OneSAF Objective System (Zacharias, MacMillan, & Van Hemel, 2008).

2004). While the OneSAF Testbed Baseline allowed for the modification of weather, environmental factors (such as temperature) did not impact the modeling of the commander-centric approach (Juarez-Espinosa & Gonzales, 2004).

d.   *Combined Arms Analysis Tool for the 21st Century*

In his discussion of the Combined Arms Analysis Tool for the 21st Century (COMBATXXI), Blais (2016) describes the discrete-event military analysis simulation as needing to provide better human behavior models that mimic actual human behavior instead of the behavior of unmanned systems. COMBATXXI does not provide realistic enough human behaviors compared to unmanned system behaviors, and an overarching challenge exists to supply human behavior models with the fidelity to include more subtle properties such as fatigue (Blais, 2016). Interestingly enough, COMBATXXI actually contains the software infrastructure to model human physiology, and this comprises an area of opportunity for creating better human behavior models that could include non-visual properties of human behavior (e.g. "fatigue, hunger, and thirst") (Blais, 2010, p. 10).

e.   *Infantry Warrior Simulation*

IWARS provides a constructive analysis simulation capability for infantry operations at the small unit level (Schleper et al., 2010; Ubink et al., 2008). IWARS focuses on analyzing the interplay of Soldiers operating on the battlefield with their associated equipment, and contains the capability of modeling physiological details such as agent heart rates (Edwards, Grodevant, Lee, & Peralta, 2007). IWARS exists for the purposes of analyzing Soldier and equipment effectiveness, and to assist with the acquisitions process of infantry equipment (Borgman, 2007; Edwards et al., 2007; Reilly, Harper, & Marotta, 2007). The use of IWARS also extends to assessment of "new tactics, techniques, and procedures (TTPs)" (Reilly et al., 2007, p. 267). Borgman (2007) notes an important balance to be maintained between human behaviors and physics models within IWARS that results in some sacrifice of behavioral realism since IWARS is not intended as a virtual reality simulation.

IWARS has been linked to a system called Modeling Individual Differences and Stressors (MINDS) Behavior Moderator Engine that worked within a cognitive

architecture called Situation Awareness Model for Pilot-in-the-Loop Evaluation (SAMPLE) in order to "model infantry squad leader decision making" (Zacharias et al., 2008, p. 74). Zacharias, MacMillan, and Van Hemel (2008) offer as a possible example the idea that MINDS can help facilitate modeling the "effect of fatigue level on perception" by serving as a "plug-in for other cognitive architectures…as a means for generating personality- or stress-based moderators that can moderate structures or parameters of the target cognitive architecture" (p. 74). In sum, Reilly, Harper, and Marotta (2007) offer that MINDS helps with modeling the "effects of behavior moderators…on perception, situation assessment, and decision making" (p. 268). The authors provide a use case of their system (via a federation of IWARS, MINDS, and SAMPLE) in an urban room clearing simulation where a team leader must analyze a situation in the room and conduct threat assessments of individuals inside based on things like the distance from an agent to a possible threat, as well as if the threat is carrying a weapon. Perception with respect to an IED danger area in the environment forms a key part of this thesis, but we assume the IED danger area location as a threat *a priori* rather than in an evaluative manner as in Reilly et al.'s (2007) room clearing simulations.

Eldridge et al. (2012) provide a description of their work on the Soldier Load and Speed-Regulation model for linking Soldier's load weight, Soldier metabolism, and resultant Soldier speed in order to measure the overall time needed to perform operational tasks. This work focuses largely on physiological effects, as opposed to cognitive effects of load weight in relation to human performance. Eldridge et al. (2012) input their model into IWARS ostensibly as a means for assessing the model's interaction with other factors such as various environmental conditions.[9] Incorporating their model into IWARS is an important example of investigating whether or not "encumbrances," such as heat stress or carrying weight, and their "effect…on task performance" can be modeled within a constructive simulation (Eldridge et al., 2012, p. 1).

Woodill, Barbier, and Fiamingo (2010) conducted an experimental effort using IWARS that sought to analyze an appropriate squad size based upon conducting urban

---

[9] The authors report that they built their model partly upon Pandolf et al.'s (1977) model.

combat operations in a hot, dry environment of over 45º C (113ºF). IWARS served as the experimental test bed, which allowed the authors to model the environmental temperature, Soldier equipment, ambient air temperature, and the activity of the Soldier and the resultant effects on core body temperature (Woodill et al., 2010). Woodill et al. (2010) point out that environmental and climactic conditions—specifically mentioning ambient air temperature's role in inducing heat stress within Soldiers—can play a part in affecting a Soldier's performance and result in behaviors such as confusion or errors in target acquisition. The authors describe how the experiment took place under three different experimental conditions: one where heat stress via the core body temperature did not cause the Soldier to stop operating, and two different conditions in which the Soldier's core body temperature exceeding the temperature setting would cause the Soldier to stop operating (the temperature settings used were 38.5ºC/101.3°F and 37.2ºC/98.96°F). Notably, the authors do not provide an explanation for why these temperatures served as their chosen settings for thresholds, and define this as a limitation of their study and outline varying the threshold values as future work. The study by Woodill et al. (2010) proves important for its constructive implementation of a combat scenario implementing heat's impact on human performance, and its calls to experiment in the future with different climates and varying Soldier performance decrement beyond simply ceasing to perform after a Soldier's core body temperature reaches a certain point.

f.   MAGTF Tactical Warfare Simulation

The Marine Corps utilizes the Marine Corps Air-Ground Task Force (MAGTF) Tactical Warfare Simulation (MTWS) constructive simulation with aggregate unit resolution in order to train Marine Corps staff members (Program Manager Training Systems, 2016). Interestingly, MTWS has the capability to modify a unit's fatigue level based on multiple factors to include temperature and humidity, and can also influence perception by affecting a unit's visual target detection ability based upon weather precipitation (MTWS Program: Cole Engineering Services, Inc., 2016). Ultimately, the environmental temperature plays no part in affecting unit path finding within MTWS.

3.   Summary

As a final point of analysis, we experimented in VBS 3 with a CGF AI's behavior in navigating to a destination waypoint on the opposite side of a danger area imposed by an actual emplaced IED. The behavior of the CGF agent in certain scenarios, although exhibiting a tactical posture, allowed the agent to get closer to the IED location than one would consider tactically sound. At other times, the CGF agent did not move from its start location, but instead changed its position between the kneeling and prone repeatedly. We did notice instances where the initial distance from the IED and the destination waypoint affected how far the agent offset its path around the threat. Larger spacing allowed the agent to have better tactical offset as it moved around the IED, while a scenario with smaller spacing contributed to the agent having poor offset from the IED. Ultimately, the most important takeaway from experimentation with VBS 3 is that the environment's temperature plays no role in affecting the cognitive thought process and decision-making of the CGF agent in its pathfinding.

The preceding offers a general understanding of the basic factors of human behavior that must play a role in military simulations. Furthermore, we have delved into some of the capabilities offered by FPS virtual simulations such as the VBS series. We have also reviewed relevant DOD simulations and their environmental modeling and pathfinding capabilities. Next, we take a more thorough dive into the subject areas of attention and vigilance, as well as simulation pathfinding techniques.

THIS PAGE INTENTIONALLY LEFT BLANK.

II. HUMAN PERFORMANCE MODELING

A. HUMAN PERFORMANCE AND THE ENVIRONMENT

A review of the relevant literature describing a human information processing model for human decision-making, as well as a review of attention, vigilance, and environmental impacts on human performance follows. This review will provide context to the cognitive processes at play in the modeled CGF agent put forth in this thesis.

1. Cognitive and Behavioral Factors

Wickens et al. (2013) offers an information processing model that will serve as a useful construct for demonstrating modifications to the decision-making process caused by environmental heat.

a. Model of Human Information Processing

The Human Information Processing (HIP) Stage Model (Figure 3) offers a simplified construct for modeling how humans process information from the environment and then translate this information into action (Wickens et al., 2013).
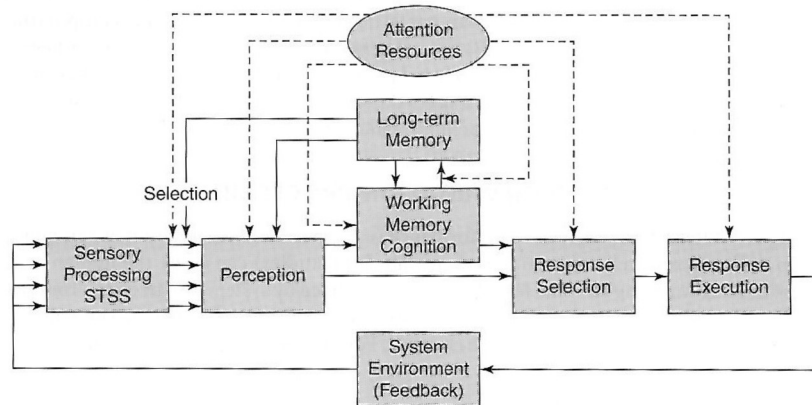
Figure 3.   Human Information Processing Stage Model.
Source: Wickens et al. (2013).

The HIP model divides into multiple components that consist of: sensing the environment, synthesizing the information from those senses into perception of the environment, combining the information that one perceives with working and long-term memory to give sensory information meaning and incorporate continuously incoming information, synthesizing all of the relevant inputs up to that point, and finally making and executing a decision (Wickens et al., 2013). Referencing Rasmussen and Rouse, Wickens et al. describes the environmental sensory information that gets "perceived" as becoming the "basis of an understanding, awareness, or assessment of 'the situation' confronting the decision maker…a process that is sometimes labeled *diagnosis*" (as cited in Wickens et al., 2013, p. 248). Wickens et al. (2013) go on to state that "this diagnosis or assessment is based upon information provided from two sources, the external cues filtered by selective attention…and long-term memory" (p. 248). In a closely related concept, working memory, theorized as existing within the HIP Stage model between the perception stage and decision-making stage, serves as a storage point to use for the means of thinking about perceived information—essentially the "cognition" stage (Wickens et al., 2013, p. 5). Additionally, the authors suggest that information processing can begin and end at any point throughout the model. Cautiously, the authors point to Wickens and Carswell as they describe the model seen in Figure 3 as a "useful framework" that "should not be taken too literally" (as cited in Wickens et al., 2013, p. 5).

Several components of the HIP Stage Model prove especially important for further analysis in this thesis. The delineation of sensing information from actually perceiving is particularly important in that: "Sensation is not perception, and of this large array of sensory information [environmental information] only a smaller amount may be actually *perceived*" (Wickens et al., 2013, p. 4). The authors also note that perception has links to lessons learned from long-term memory that plays a crucial role in "determining the *meaning*" of this sensory information (p. 4).

Attention serves as another component of the HIP Stage Model that deserves further review in the context of this effort. Wickens et al. (2013), referencing Wickens and McCarley, point out that two functions of attention exist within the process model (as cited in Wickens et al., 2013). These functions are attention as an information "*filter*,"

and attention as a source of "*fuel*" for governing the model (Wickens et al., 2013, p. 5). The information filter role means that attention serves as a selection mechanism that allows individuals to key in on important information (believed to be more important by the human processor, i.e., having "higher perceived value"), while discarding other elements of information (Wickens et al., 2013, p. 248). Wickens et al. (2013) describe how the role of attention as an information filter translates directly to the cognitive property of "selective attention," and point to long-term memory as the basis for selective attention (p. 49). The idea that attention could serve as a source of fuel corresponds to attention serving as a mechanism for supplying "mental resources" to key parts of the model; for example, the authors offer the analogy of a driver on a foggy night requiring more attentional fuel supplied to the perception stage of the process model than they would under normal conditions (p. 5). The suggestion of differing "fuel" requirements due to environmental factors serves as inspiration for a scenario for this thesis that envisions a hot, humid patrolling environment that would place more demands on the perceptual portion of a Marine's information processing model.

Another important factor concerning the HIP model focuses on the draining of attentional resources, resulting in negative affect on the performance of multi-tasking (Wickens et al., 2013). In theory, the authors describe how drained attention could lead to mistakes made in later portions of the model for other "concurrent" tasks, such as making a decision (p. 5). This approach suggests the information process can be divided, per Mosier and Fischer, into a "front end" (essentially the sensory and perceptual portions) and "back end" (referring to the actual decision made by the human) (as cited in Wickens et al., 2013, p. 248). Making a decision at the end of the information process brings about a risk estimation process with respect to what could come about as a result of different "decision options" (Wickens et al., 2013, p. 248). Wickens et al. (2013) point out that human decision making can fail early in the process (i.e., the "situation assessment" stage) for a multitude of reasons (p. 248). The authors, referencing Hoffman, Crandall, and Shadbolt, argue that discerning the "overall distinction between front-end and back-end processes" proves "critical to understanding decision failures" (as cited in Wickens et al., 2013, p. 248). The authors point to the work of Weigmann, Goh, and O'Hare, in

describing the importance of this distinction, "Very different solutions may be applied to remedy environments where decisions fail because of poor information and situation assessment, compared to those when failures result from inappropriate (e.g., too risky) choices in the face of a well-diagnosed situation" (as cited in Wickens et al., 2013, p. 248).

b.   *Attention and Vigilance*

In order to provide a mechanism for modulating attention and vigilance within a CGF agent's human information processing behavior, a more thorough review of relevant definitions of attention and vigilance proves necessary. Wickens et al. (2013) provide simplified definitions of a few relevant types of attention—specifically selective, focused, divided and sustained attention. The authors illustrate how sustained attention correlates strongly with vigilance (discussed below), divided attention describes paying attention between multiple information sources, and focused attention describes honing in on key pieces of information and ignoring distraction. Pew and Mavor (1998) describe divided attention and multi-tasking as highly relevant to military combat operations. The authors even note an example of an infantryman having to path find at the same time as completing other combat tasks. Wickens et al.'s (2013) focus on selective visual attention proves quite relevant when considering certain military operations such as patrolling—especially when one sees that patrolling incorporates many of the tasks that rely on selective visual attention. The authors describe some of these relevant selective visual attention types of tasks: "*general orientation and scene scanning*," as well as "*noticing*"—essentially maintaining a look out for and "responding to…somewhat unexpected events" (p. 50). Selective visual attention also includes "*searching* for specific, usually pre-defined targets" (p. 50).

Chun, Golomb, and Turk-Browne (2011) describe attention as consisting of multiple characteristics to include limited capacity, selection, modulation, and vigilance. The limited capacity and selection characteristics roughly correspond to Wickens et al.'s (2013) description of how not all sensory information from the environment can get perceived by an individual, and the idea of attention as a perceptual filter (Chun et al.,

2011). Chun et al. (2011) extends the idea of selection by describing it essentially as decision-making between possible options. The concept of modulation corresponds most strongly with Wickens et al.'s (2013) description of attention as a fuel within the HIP model. Chun et al. (2011) describe modulation as the degree to which "attention determines how well the target information is processed," further, the speed and accuracy with which tasks or responses get performed serve as key additional details (p. 75). Similarly, vigilance is described as being related to the modulation component of attention, but focused on the "ability to sustain…attention over extended periods of time" (p. 76). As a matter of practicality for this thesis, the concept of attention manifests itself in a CGF model because there often exists an inordinate amount of information in an operational environment to which an individual must attend. Similarly, Chun et al.— referencing the work of Treisman and Gelade, and Wolfe and Horowitz—point out that "in natural contexts, observers typically search for a target amid many other competing stimuli" (as cited in Chun et al., 2011, p. 87). The authors eloquently summarize these challenges in a manner that could apply to military ground combatants, making attention a relevant human characteristic to model in a CGF: "The problem, that there is too much information to process; the solution, to select and modulate information most relevant for behavior; and the challenge, to sustain vigilance" (p. 76).

A handful of studies designed to measure vigilance prove useful when describing vigilance as a cognitive process. Notably, Szalma, Teo, Hancock, and Murphy (2011) examined the relationship between vigilance and tactical tasks and identify the importance of vigilance training for detecting IEDs, even proposing video game-based approaches for doing so. In another study, vigilance performance in terms of percent correct detections of stimuli (auditory, visual, and tactile) suffered from load carrying during completion of an obstacle course (Mahoney, Hirsch, Hasselquist, Lesher, & Lieberman, 2007).

c. *Attention, Working Memory, and Visual Search*

Chun et al. (2011) splits the classification of attention into two categories— internal and external (generally perceptual) attention—where internal attention deals

primarily with cognitive characteristics such as the general areas of working memory, long-term memory, and response selection.[10] The authors further hypothesize that working memory sits "at the interface between internal attention and external attention," but places it within internal attention "because it operates over internal representations (of what is no longer externally available)" (p. 85). Working memory limitations with respect to vision consists of the ability to store approximately four objects (Chun et al., 2011). This limitation could prove important within the context of this thesis when considering environmental impacts on human cognitive performance. Visual search constitutes a task essentially related to selective attention that proves highly applicable to the means with which one can model an environmentally impacted CGF agent (Wickens et al., 2013). Simply put, visual search results from attempting to find a "target" utilizing one's eyes as one applies their selective attention in the given search area, and the target of a search task receives definition prior to the task taking place (Wickens et al., 2013, p. 56). Chun et al. (2011) note that visual search tasks prove highly related to human mistakes (consider failed security baggage screening among other examples). However, the authors (summarizing the works of Bar; Chun; Torralba, Oliva, Castelhano, and Henderson) note that visual search can be aided "when targets appear in predictable locations, cued by background context and past experience" (as cited in Chun et al., 2011, p. 87).

A synthesis exists resulting from the work of Oh and Kim, and Woodman and Luck, concerning the inter-relation of visual search and working memory in terms of humans executing visual search and spatial working memory tasks at the same time that can result in reduced search efficiency—meaning that both separate tasks exhibit a "shared capacity" (as cited in Chun et al., 2011, p. 85–86). Similarly, vigilance tasks requiring a considerable amount of mental resources can prove vulnerable to interference when performed simultaneously with other tasks (Wickens et al., 2013). Further, Caggiano and Parasuraman found an increased "vigilance decrement" can result when a

---

[10] Wickens et al.'s (2013) information processing model incorporates a "response selection" component— what we have subsequently referred to as the decision-making point within the model (p. 4).

simultaneously executed vigilance task and a separate task both require resources from spatial working memory (as cited in Wickens et al., 2013, p. 28).

Visual search serves as a crucial component of numerous military applications, especially infantry patrolling tasks, where the infantryman must scan their field of vision for enemy targets, IED threats, danger areas, and a multitude of other relevant information. Often times, training and past enemy tactics will guide infantry personnel to expect threats from certain places, such as an IED placed under a road in a culvert (the IED danger area used in our model). Accounting for the interconnectedness of attention, vigilance, visual search, working memory, and other tasks as we have outlined from the literature would lead to robust and interesting manipulations within CGF agents executing similar tasks within a FPS simulation. As previously alluded to, the purpose of this thesis is to model the interplay of heat, attention, and vigilance while patrolling, ultimately showing how heat can negatively affect attention and vigilance such that proper dispersion from threats, or danger areas, in a Marine's chosen path declines. Due to the complexity involved with fully incorporating working memory and visual search into a potential CGF model, we will first assume that working memory performs its cognitive role as described within Wickens et al.'s (2013) model. Next, we will also assume that the agent within our model executes visual search, but that they will be aware of the IED danger area *a priori*.

2. Environmental Factors

Environment impacts of heat have been shown to have a significant impact on cognitive performance. In their 2007 meta-analysis, Hancock, Ross and Szalma completed an extensive statistical analysis of fifty-seven studies derived from a large body of literature where the impacts of environmental heat and cold on human performance were measured. Among the key takeaways from this work centers on its finding that "performance under thermal stressors proved on average to be approximately one third of a standard deviation or about 11% worse than performance at a comparative thermoneutral temperature" (Hancock, Ross, & Szalma, 2007, p. 857). This finding incorporates both hot and cold temperature stressors; however, the mean effect size due

to heat ($\bar{g}$ = -0.29) alone proved only slightly lower than the mean effect size found for the overall meta-analysis including both hot and cold stressors ($\bar{g}$ = -0.34) (Hancock et al., 2007).[11] A key requirement of the Hancock et al. (2007) meta-analysis included the need for each study to have a room temperature control group present for comparison to the experimental condition, thus negative numbers for mean effect sizes signify decremented performance in the experimental condition.[12] An earlier meta-analysis performed by Pilcher, Nadler, and Busch (2002) includes an additional factor for analysis concerning how pre-task exposure to temperature (especially for greater than one hour) results in negative performance of tasks in the heat. The inclusion of pre-task exposure approach draws a parallel to a combat patrolling environment in which military personnel may live and be acclimatized to hot surroundings prior to executing a patrol.

Hancock et al.'s (2007) meta-analysis also provides an interesting breakdown of the various ways in which they analyzed their data. For example, a few of the groups that the studies were classified into included ones that split the studies into a category for heat at a level of 85º Fahrenheit (F) Effective Temperature (ET) (equivalent to 87.4ºF Wet Bulb Globe Temperature (WBGT)), as well as a breakdown along the lines of perceptual, cognitive, and psychomotor tasks. Notably, Hancock et al. (2007) required that enough information existed in a study to allow for calculating the WBGT in order for the study to make it into the meta-analysis. Other important category breakdowns offered in the Hancock et al. (2007) meta-analysis, related to time of task exposure for combined heat and cold studies into one hour blocks from one to three hours (0-1 hour, 1–2 hours, 2–3 hours and greater than 3 hours), categories to measure temperature's impacts on speed and accuracy of task completion, and breakouts into multiple temperature ranges (78.3ºF ET – 85ºF ET, 85 - 95.4º F ET, and greater than 95.4ºF ET prove most important for this thesis). The higher ranges of 85–95.4ºF ET, and over 95.4ºF ET exhibited the largest

---

[11] See Hancock et al. (2007) for a detailed description of the statistical methods used for calculating effect sizes, as well as more explanation of outlier analyses—a purported extension of this work over previous work done by Pilcher, Nadler and Busch (2002).

[12] Pilcher et al. (2002) take a similar approach.

mean effect sizes in favor of heat having an effect on performance (-0.40 and -0.39 respectively) (Hancock et al., 2007).

Hancock et al.'s (2007) meta-analysis argues for a 85ºF ET "threshold" as the point of demarcation for analysis due to the fact that the body begins to store heat without any resistance, and the body's core temperature begins to rise at this temperature (p. 860). Hancock et al. (2007) refers to multiple studies as support for this threshold. There is more to this finding than just the storage of heat in the body. The authors report how multiple variables aside from heat can contribute to performance decrements: "Mean performance change alone is insufficient to represent a full picture of what is going on as stress increases and, instead, that increasing variability in performance scores represents a crucial indicator" (p. 860). This delineation proves important due to the similar mean effect sizes between the greater than and less than 85ºF ET categories ($\bar{g}$ = -0.25 and -0.27, respectively), while their effect size variances differ greatly (0.67 and 0.24, respectively) (Hancock et al., 2007). The authors believe that multiple confounding variables such as exposure time to the temperature, type of task, and "other environmental factors" likely contribute to this result (p. 860).

As previously mentioned, Hancock et al.'s (2007) meta-analysis includes an examination of duration of exposure time in one hour blocks from less than an hour to greater than three hours, but do not classify these by hot or cold temperatures. The authors describe how the greatest increase in performance decrement occurs as one increases in time from the 1–2 hour mark and then to the 2–3 hour mark (having the greatest decrement). They also point out that although the greater than three hour zone and less than one hour zone had similar low effect sizes for human performance of tasks this constituted a statistical "artifact," and that the greater than three hour zone still proves bad for human performance and is generally intolerable (p. 863). The authors mention that their meta-analysis could not find enough studies that incorporated acclimatization—a topic that would have implications for analyzing heat exposure duration results—and point to it as an area for future work.

Hancock et al. (2007) do provide some categorical break out of temperature depending upon whether the tasks occurred for greater than two hours, or less than two

hours. In this study, they found that less than two hours of task performance in greater than 85ºF ET proved worse for performance than if performed for longer than two hours, and allude to "acclimation" as a potential reason for this difference (p. 867).

Another review of the extant literature revealed numerous studies that support the 85º demarcation line (although they focus on 85ºF WBGT rather than ET)—especially as it relates to visual vigilance tasks and vigilance (Johnson & Kobrick, 2001). Johnson and Kobrick (2001) conclude from their analysis of several studies that 85ºF serves as the best temperature for the performance of visual vigilance tasks. In this investigation, the authors focus specifically on tasks such as watch standing or surveillance, stating that heat serves to primarily effect the visual portion of these tasks. From their review of studies, the authors also attribute boring, simple tasks (or "low arousal" tasks) in conjunction with heat as another aspect causing hindered vigilance performance (p. 154). The authors report on a National Institute for Occupation Safety and Health criteria report that 85ºF WBGT serves as the upper limit for exposure to the heat if one desires "unimpaired mental performance" for a human executing a task for the length of time of "two hours or longer" (as cited in Johnson & Kobrick, 2011, p. 137). Furthermore, in reference to the same source, the authors describe how one hour of task performance when temperatures exist around 109ºF WBGT serves as another limit for acceptable performance. Johnson and Kobrick (2001) also provide an extensive treatment of multiple studies pertaining to acclimatization and human performance in heat that prove interesting, but beyond the scope of this thesis.

Hancock et al.'s (2007) breakout of the studies for all temperatures according to where they fall within an information processing construct is important in that they found the greatest effect of heat stressors on tasks relating to perception, then psychomotor response, and lastly cognitive types of tasks. Interestingly when the studies were broken out according to greater or less than 85ºF ET, the authors could not offer a conclusive analysis with respect to perception studies grouped in the temperatures below 85ºF ET category due to too few effect sizes available in the meta-analysis study sample fitting those parameters. However, Pilcher et al.'s (2002) meta-analysis did find a more degraded degree of performance in the studies grouped as attention and perception tasks

for temperatures over 80ºF WBGT, as compared to the level of negatively affected performance found in attention and perceptual task studies grouped into the category where the temperature stood at 65ºF or less.

Fortunately, enough studies existed for the Hancock et al. (2007) meta-analysis to analyze the 85ºF ET break down according to task completion accuracy and reaction time, yielding results where response time increases above 85ºF ET. Interestingly, when outliers were removed from their study they found that "accuracy is impaired more above 85ºF (29.4ºC) ET, whereas impairment of speed occurs more at temperatures below 85ºF (29.4ºC) ET" (p. 867). Johnson and Kobrick (2001) comment on the difficulty of analyzing heat's effects on reaction time due to differences in the approaches and study conditions in the studies they looked at, ultimately concluding that the "resulting findings tend to be contradictory" (p. 138). The authors also identify greater reaction speed variability as taking place when temperatures are high or in complex task performance. Similarly, Hancock et al. (2007) found a greater amount of variability for response time performance as compared to accuracy for thermal stressors in general, likely because of moderator variables more apparent in the response time category. Ultimately, enough of a relationship exists in the analysis to see that heat interacts with speed and accuracy in task performance—especially when Hancock et al. (2007) break down their analysis to compare heat's effects on perceptual accuracy and perceptual response time which showed significant decreases in accuracy and response time.

The importance of the conclusions offered by Hancock et al.'s (2007) work centers around multiple factors, particularly: "The magnitude, and in some cases the direction, of the effect of thermal conditions on performance depended on particular combinations of exposure range (i.e., heat/cold), task type, performance measure, and the duration and intensity of exposure" (p. 870). The authors also offer that their findings support an "extended-U" curve (see Figure 5) over the Yerkes-Dodson (Yerkes & Dodson, 1908) "inverted U" curve (see Figure 4) that compares arousal to human performance; the extended-U curve offers a broader range of "relatively stable" human performance at a higher level bookended by significant performance drop offs as physiological and psychological stress levels increase (Hancock et al., 2007, p. 871).

Hancock and Warm (1989) and Hancock and Vasmatzidis (2003) offer more in depth descriptions of this extended-U curve, also called the Maximal Adaptability Model. The key takeaway from this model for our purposes of this thesis comes from Hancock and Vasmatzidis (2003) who reference the work of Kahneman in stating a model assumption "that heat exerts its detrimental effects on performance by competing for and eventually draining attentional resources" (as cited in Hancock & Vasmatzidis, 2003, p. 367). This idea appears quite similar to our discussion of Wickens et al.'s (2013) discussion of attention as a fuel.
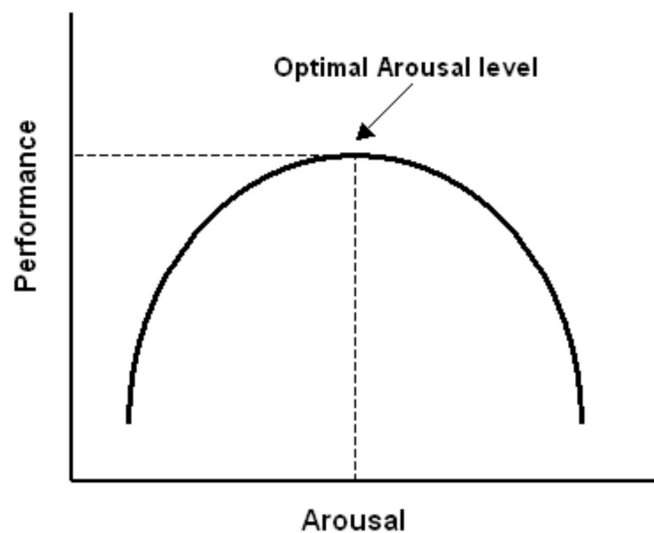

Figure 4.   Yerkes-Dodson Inverted-U Curve.
Source: Hancock and Vasmatzidis (2003).
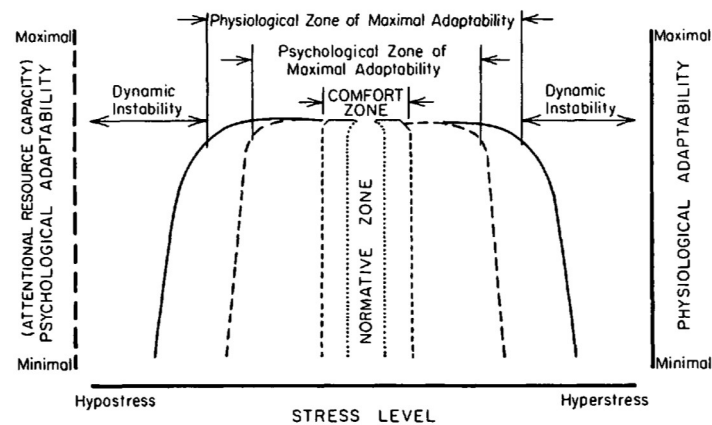
Figure 5.　Hancock and Warm's Maximal Adaptability Model.
Source: Hancock and Warm (1989).

B.　BUILDING HUMAN PERFORMANCE MODELS

The following section details frameworks and techniques for modeling human performance. As previously mentioned, the realism of human behavior models is a key factor to this investigation. The overarching call for creating more unpredictable and adaptive models of human behavior is not necessarily a new requirement (North Atlantic Treaty Organization (NATO), 2009; Pew & Mavor, 1998; Wellbrink, 2003). Further, the NATO Task Group 128 report, referencing the work of a previous NATO report from Dompke, describes a shortage of necessary behavior moderators resulting in "brittle, non-adaptive, and predictable" models (as cited in North Atlantic Treaty Organization (NATO), 2009, p. 1–3).

The focus of NATO Task Group 128's report centers on constructive simulation implementations of human behavior, and provides extensive description on a range of topics including performance shaping functions and multiple resource models of human information processing (similar to Wickens et al.'s (2013) model, albeit more simplified) (North Atlantic Treaty Organization (NATO), 2009). The report also describes modeling efforts in cognitive architectures. Ultimately, the Task Group 128 report finds that "modeling of post-receptive perceptual processes such as attention…is perhaps the best developed aspect of perceptual modeling by the HBR [Human Behavior Representation] community" (p. 4–7).

We benefit from a number of NATO Task Group 128 observations of techniques used within the field of constructive simulation in this thesis. One technique stems from the lack of ability in virtual environment simulations to provide detailed models of environmental stimuli, requiring effective shortcuts where one supplies the details (i.e., position or identifiers) about the environmental stimulus directly to entities by embedding the data within the entity's code (North Atlantic Treaty Organization (NATO), 2009). Another observation from the report comes from the field of error modeling and stochastic discrete event simulations, where the report relates how error modeling can be derived by sampling from the "extremes of the distribution of the task times or accuracies" (p. 8–10). Finally, the NATO Task Group 128 report offers that an alternative to performance shaping functions for modeling the effects of stressors involves using a heuristic to model an environmental stressor's impact on human performance.

1.   Waypoint Graphs and A* Search Algorithm

Military virtual training simulations and commercial FPS video games often utilize an A* pathfinding search in conjunction with waypoint graphs in order to support the movement of an agent within the game's environment. The A* search algorithm, in conjunction with a mapping of a game environment via a waypoint graph representation, serves as a mechanism for allowing agents to navigate paths within a virtual game or simulation environment (Tozour, 2003). Russell and Norvig (2010) and Stout (2000) provides a concise overview of the A* search algorithm. Stout (2002) also offers a few approaches to dividing an environment into a search space.

a.   A* Search

Russell and Norvig (2010) describe A* search as an algorithm that "evaluates nodes," such as those in a simulation environment's waypoint graph, "by combining $g(n)$, the cost to reach the node, and $h(n)$, the cost to get from the node to the goal" (p. 93). The combination of these two costs results in the "estimated cost of the cheapest solution through [node] n," written as "$f(n)$" (p. 93). Stout (2000) terms $g(n)$ as the "CostFromStart(X)" (where "X" represents the node under evaluation), and $h(n)$ as the "CostToGoal(X)" in pseudocode (p. 255). The cost from the start usually consists of total

"distance traveled" to that node, but "other factors can be added into this function, such as penalties for passing through undesirable areas" (Stout, 2000, p. 259). The cost $h(n)$ is also called a "heuristic" cost (Stout, 2000, p. 255). Similar to Russell and Norvig's (2010) naming conventions, this thesis will refer to a node's cost from the start as the "g" cost, to the node's heuristic cost as the "h" cost, and to the node's combined cost as the "f" cost (see Equation 1.1).

$$f = g + h \tag{1.1}$$

Russell and Norvig (2010) provide an outline for how heuristic costs must meet the principle of "*consistency*" via a triangle inequality in order to be optimal during A* searches of graphs (p. 95).[13] A heuristic cost must also prove "admissible" in that it "*never overestimates* the cost to reach the goal*" (Russell & Norvig, 2010, p. 94). Both Russell and Norvig (2010), Rabin (2000b), and Stout (2000) provide some commentary on the value of choosing appropriate heuristics for speeding up searches with the downside being paths that may no longer prove to be the most optimal due to overestimating heuristic values. Stout (2000) describes another approach for decreasing search time via utilization of a "*Closed*" list; essentially the algorithm ignores previously explored locations (p. 255). Stout (2000), Rabin (2000b), and Darken (2016) provide inspiration for this thesis to achieve our desired agent behavior in using the overestimation technique in manipulating the heuristic cost of waypoints, as well as manipulating the additive cost of the distance traveled to reach the current node.

b.   *Waypoint Graphs*

Tozour (2003) provides an overview of different approaches to constructing a waypoint graph within a game environment, as well as the fundamental definition that we will use going forward. Importantly, Tozour identifies that "fundamentally, all search spaces are *graphs*: they all consist of some number of atomic units of navigation, or *nodes*, and some number of connections, or *edges*, between pairs of nodes" (p. 85).

---

[13] Russell and Norvig (2010) define an "*optimal solution*" as having the "lowest path cost among all solutions" (p. 68).

Essentially, the fundamental way that agents automatically navigate around a virtual environment entails utilizing a search algorithm, such as A*, to search through a waypoint graph, called a "search space," looking for the "*optimal* path" to follow (Tozour, 2003, p. 85–86). Tozour (2003) also points out that the size of a search space has implications for search speed and memory usage. Furthermore, the author describes a key nuance in the definition of what constitutes an optimal path. The textbook definition offered by the author of optimal paths, "That is, the least expensive path from [point] A to [point] B," at times conflicts with the reality "that the optimal path is not always the shortest path" (p. 86). To illustrate this, Tozour offers an example where the faster route for an agent consists of a longer curved road avoiding a swamp rather than the slow, straight through path.

Tozour (2003) highlights two related approaches to creating search spaces: grids of squares for two-dimensional game environments, and waypoint graphs for three-dimensional game environments.[14] The author describes how waypoint graphs allow for a game designer to put waypoints anywhere within the game and can prove useful for navigation in a number of indoor and outdoor environments. A technique that allows for more realistic-looking paths in both grid and waypoint graph approaches is called "string-pulling," which essentially ignores unnecessary path nodes if two nodes can be "directly" connected (Tozour, 2003, p. 89). Tozour (2003) mentions another method known as Catmull-Rom splines, and describes how these essentially provide a "smooth, curved path" between two points (p. 89).[15]

2.   Advanced Modifications to A* Search

An advanced A* search implementation for creating more realistic path finding behavior of agents based upon the tactical situation is put forth by van der Sterren (2002). The author describes how his approach—"*tactical path-finding*"—alters the A* algorithm's heuristic to account for moving through enemy fire and observation areas, as

---

[14] See Tozour (2003) for a number of other search space approaches, including navigation meshes.

[15] Rabin (2000a) describes how to implement Catmull-Rom splines in detail.

well as the ability for the enemy to aim effectively, thus enabling an agent to move realistically between positions of cover and concealment (p. 294). Further, the author relates how he does so via a weighting approach for each of the tactical criteria within the desired behavior of covered and concealed path finding. The author points to Reece et al. (2000) who utilized a similar approach in their ModSAF implementation, arguing that both A* implementations similarly "balance short travel time versus visiting risky locations, based on the weights for each type of cost" (as cited in van der Sterren, 2002, p. 296).

Van der Sterren (2002) provides several other worthy considerations for enhancing the A* algorithm to make it more realistic. Some of these include accounting for the "distance to threat and the threat's weapon," since these both "might also have a large influence on the risk of being in the line-of-fire" (p. 301). The author also advocates intelligently applying his techniques to go beyond just "avoiding known and suspected hostile lines-of-fire," but to use them to "avoid weak combat positions" (p. 301). The author proposes using his techniques in conjunction with those of van der Sterren (2001) to assess more cost to weak positions, thereby encouraging agent paths that "will try to avoid them" (p. 301).

Van der Sterren's (2001) approach to extracting information from waypoint graphs allow for more dynamic artificial intelligence of CGF entities since one would prefer CGF "capable of automatically annotating" a game environment, rather than resorting to "manually editing dozens of waypoints" with specific properties (p. 2). Harder and Darken (2016) provide an illustration of a non-manual approach: "Given fixed enemy positions, we can perform a preprocessing step to tag all locations in the navigation graph with their valid targets" (p. 7). Van der Sterren (2001) illustrates his approach via an example where an entity charged with executing a task as a sniper analyzes the game's terrain via a mathematical formula that incorporates terrain and sniper position properties to choose the best "sniping spots" (p. 1).

The approach utilized by van der Sterren (2001) demonstrates the possibilities for applying the model to the problem of dynamically assessing danger area locations, or more generally positions occupied by the enemy. A danger area approach to terrain

reasoning could re-purpose van der Sterren's (2001) sniper equation heuristic that assesses a sniper position's ability to over watch locations with heavy enemy traffic, to a heuristic that classifies enemy over watched terrain at choke points highly frequented by heavy friendly traffic (we propose a bridge over a creek culvert as an example) as a danger area. Borkman, Verdesca, and Watkins (2012) implement a similar idea in a heuristic for an agent's route that determines danger areas based on what positions the enemy will likely use to watch the agent's route. Lidén (2002) offers a paper similar in nature to that of van der Sterren (2001) in utilizing waypoints to derive meaning from the game's tactical environment, but different in that Lidén stores such information in bit format. Furthermore, Lidén (2002) offers a technique for assessing a waypoint graph for "pinch points," essentially choke points, within an urban environment (p. 7).

Straatman, van der Sterren, and Beij (2005) conduct research in a similar vein, offering their technique of "dynamic procedural combat tactics" utilizing "position evaluation functions" for terrain assessment within a commercial video game known as *Killzone* (p. 1). The authors describe inflexible tagging as providing limited utility for dynamic situations, or where the "environment is likely to change" (p. 2). Their effort proposes position evaluation functions as the alternative, and illustrate how to use these for picking positions for agents and for promoting tactical path finding for agents. Essentially, a waypoint has several tactical and environmental characteristics about itself evaluated and combined into a numeric representation given the tactical situation; this waypoint's numeric value then gets ranked against adjacent waypoint values, and the waypoint possessing the highest value serves as the position for the agent to move to (Straatman et al., 2005). The authors describe how they also apply this technique in a similar manner to van der Sterren's (2002) tactical A* approach, where path waypoints are evaluated and contribute to the A* path's cost function. Importantly, Straatman et al. (2005) offer how a position evaluation function can be modified by adjusting the weighting of its components resulting in various behaviors. For example, one could adjust the function's weights such that attacking agents prefer "out in the open positions over positions near cover," or one could have agents factor in properties such as the distance from themselves to a main threat (p. 8).

Borkman et al. (2012) describe a research effort implemented utilizing IWARS for creating AI algorithms focused on route planning and finding over watch positions. The authors demonstrate how their route planning algorithm utilizes A* search, and incorporates numerous heuristic values based upon multiple tactical considerations, called "factors," such as where the enemy or positions of cover could be (p. 21). This approach appears quite similar to many of the previously discussed techniques. A key takeaway from Borkman et al. (2012) includes their implementation of line-of-sight ray casting from a possible agent over watch position's "view point" to a pixel field (called a "raster") as a means for determining what the agent can observe (p. 29). Darken (2004) describes this method in detail, as well as a simpler method (i.e., fewer rays) called "multiple ray casting" (p. 3).

A key consideration in this thesis's creation of the implementation of a heat-influenced limited attention pathfinding algorithm entails how the CGF entity will perceive the presence of an IED threat from the culvert programmatically speaking. Abdellaoui et al. (2009) highlights that achieving realistic behavior in CGF requires that a choice exists for how entities will perceive their environment between detecting the environment via their own "sensors," or having "ground truth" fed to them automatically (p. 2–5). Wray et al. (2002) provide an example of this automatic feed of ground truth via their use of "map annotations" to allow Unreal Tournament CGF to know their location within the game environment (p. 6).

3. Limitations of A* Search

Logan and Alechina (1998) warn that aggregating cost functions in a manner similar to the approach advocated by van der Sterren (2002) can lead to an inability to determine which adjustment to a specific performance criteria led to the outcome behavior, and often means wholesale changes to each criteria's weighting in order to manipulate the weighting of only one. Another concern with multiple-criteria weighting problems, specifically in the path finding domain, is that they may require different weighting combinations per specific scenarios—essentially one set of criteria settings may not prove equally useful or applicable across all possible scenarios (Logan &

Alechina, 1998). Furthermore, the authors point out that "many real-world problems are difficult or impossible to formulate in terms of minimising (*sic*) a single criterion," and state that path planning "can be difficult to formulate…in terms of minimising a single criterion" as well (Logan & Alechina, 1998, p. 444). Stated another way by Ritter et al. (2003), "It is also far from clear that human behavior solely optimizes on a single criterion" (p. 30).

C.   THESIS MOTIVATION

The motivation for this thesis stems from my previous experience as an infantry platoon commander utilizing the DVTE system for mission rehearsals of training missions. In one particular rehearsal, the lack of realism inherent in the avatars representing myself and the key leadership of my platoon led to a lack of "buy in" to the usefulness of the simulation for training for our upcoming training mission. The avatars exhibited superhuman qualities such as covering multiple kilometers in "sprint" mode in mere minutes, rather than the hours that such a movement would require. Although such a capability proves ideal for focusing a simulation exercise on the key training objectives, rather than spending significant amounts of time on less important tasks, this quality would not prove ideal for ground combat units hoping to use the simulation for training at patrolling tasks where movement speed and relative location to fellow unit members does matter. Furthermore, this lack of realism led many of us utilizing the simulation to see it as more of a toy, and less as a training tool.

D.   HYPOTHESES

This goal of this thesis is to serve as a proof-of-concept for how virtual training simulations utilized by the U.S. Marine Corps and U.S. Army can embed more realism within the simulation to help promote more faith in simulation as a necessary training tool. As previously mentioned, the U.S. Army has taken steps to increase the realism of the portrayal of Soldier avatars within VBS 3 to include modeling the impacts of fatigue and physical fitness test performance resulting in, as a U.S. Army Training and Doctrine Command analyst points out, "Small unit leaders" receiving the "capability to understand the performance of their squad" (Lopez, 2014, paragraph 5). Similarly, this thesis

identified a gap in virtual training simulations where Marines do not receive an adequate representation of the decremented decision-making capabilities of their fellow Marines as a result of the hot, humid environments that the Marine Corps currently operates in. Essentially, Marines and Soldiers would greatly benefit from training where they could learn to recognize "when a given individual is showing appreciable changes in performance and cognition" (Lowe et al., 2007, p. S98). Pew and Mavor (1998) echo this need:

> For training…the behaviors that are important to represent realistically are those that can be observed by the other participants in the simulation, including physical movement and detection and identification of enemy forces. It is important that observable actions be based on realistic decision making…achieving realism with respect to these observable outcomes requires that the models of human behavior employed in the simulation be based on psychological, organizational, and sociological theory. (p. 1–2)

This thesis seeks to address this gap to promote a better training experience for Marines and Soldiers. In order to do so, our proposed model will test the following hypotheses:

1. *Hypothesis $1A_0$:* The mean difference between the random number generator (RNG) derived hot and mild temperature condition normal distribution samples equals zero.

2. *Hypothesis $1B_0$:* There is no difference between the mean of the respective temperature condition's RNG-generated sample normal distribution and the specified mean for the temperature condition's normal distribution.

3. *Hypothesis 2_0:* The mean difference between each temperature condition sample's IED Danger Area Radius sizes equals zero.

4. *Hypothesis 3_0:* The mean difference between each temperature condition sample's path lengths equals zero

THIS PAGE INTENTIONALLY LEFT BLANK.

III. METHODOLOGY

As discussed in Chapter II, certain temperature conditions in a given environment can deleteriously affect one's vigilance and attention to their surroundings. In order to explore improving the fidelity of autonomous agents for use in military-relevant simulations, this thesis proposes the development of an agent behavior model that represents decremented perceptual capabilities due to thermal heat factors. To test this model, we developed a scenario based on a common patrolling task. The scenario centers on a ground combatant encountering a bridge over a culvert which provides passage over a small, dry streamed—a potential IED danger area (United States Marine Corps, n.d.). The agent must navigate around the danger area to the opposite side in order to continue the patrolling mission. We incorporate recommended minimum standoff distances for personnel in the course of a unit's standard reaction to an IED as the basis for the general premise that if one senses an IED danger area in their path they will take measures to avoid the danger posed by that location by offsetting their direction of travel (United States Marine Corps, n.d.).

Based on empirical evidence provided by Hancock et al. (2007) suggesting that one can expect human performance perceptual decrements in hot environments, we make the base assumption that the agent will choose a path closer to the recommended offset from a danger area, or potential threat, as the agent patrols around the danger area in mild temperatures. Conversely, a hot temperature environment will result in agents achieving an offset less than the recommended distance due to decremented vigilance and attention to the potential threats in the environment. These assumptions guided an initial implementation within an experimental modeling environment for the A* search algorithm, as well as in the robust model implementation created for the proof-of-concept.

All development took place within the Unity game engine utilizing Microsoft Visual Studio 2015 as the underlying code development environment for scripting behaviors (Microsoft Corporation, 2016; Unity Technologies, 2016; Unity, 2017). Creation of the proof-of-concept and all statistical analysis testing took place on an Asus

Notebook UX303UB running Microsoft Windows 10 Home Edition with an Intel Core i7-6500U two-core 2.50 GHz processor with 12.0 GB of Random Access Memory (RAM).

A. DEVELOPMENT OF MODEL PROOF-OF-CONCEPT

Prior to developing the final model, the *A* Explorer* program served as a platform for providing an initial illustration of the final model's intended visualization under both a "hot" temperature condition and a "mild" temperature condition (Matthews & van der Sterren, 2002). *A* Explorer* implements the tactical path finding enhancements to the closed list form of the A* algorithm as described by van der Sterren (2002) (Matthews & van der Sterren, 2002).

Many types of IEDs require outdoor standoff distances greater than the recommended minimum standoff distance (National Ground Intelligence Center, n.d.). To mimic this, we used *A* Explorer's* enhancements to incorporate as extra costs both the threat agent's line of fire to a potential path cell and additional cost for the distance from the threat to the potential path cell (Matthews & van der Sterren, 2002). In our environmental model's case, the threat agent's line of fire cost represents the possibility of a fragmentation pattern from an exploding IED (assuming an open environment with few obstacles to blast fragmentation) reaching the potential path node, and the cost for distance of the path node from the threat will help assess the needed offset from the IED danger area. These additional costs get incorporated into the "g" portion of the cell's total cost (Matthews & van der Sterren, 2002).[16]

As shown in Figure 6 and Figure 7, cell colors within the *A* Explorer* environment correspond to additional costs assessed to that node. Each cell within the black border walls delineating the navigating agent's "operational environment" received an additional cost assignment to help facilitate the agent's navigation along a specific path to illustrate desired behavior per the temperature condition present in the

---

[16] We used the diagonal movement setting in our *A* Explorer* models and the setting for movement costs was set to assign cost on a cellular basis.

environment. In the mild temperature environment (Figure 6), the agent navigates along a path more offset from the IED danger area. In the hot environment (Figure 7), the agent's path travels much closer to the IED danger area as it travels around it. A black line connecting the start and goal cells depicts the agent's path generated by the A* algorithm.
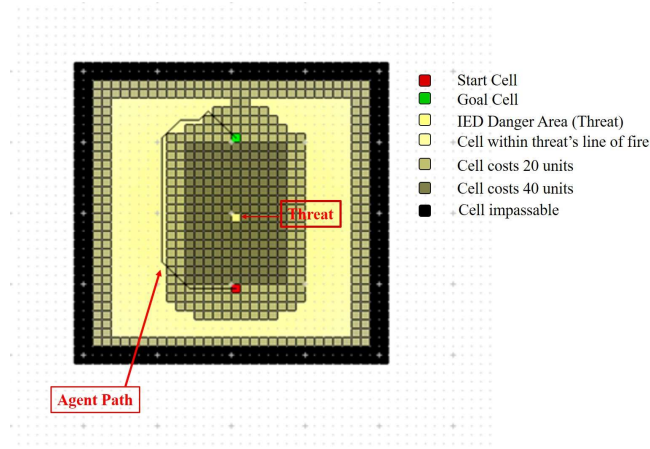

Figure 6.   Mild Temperature Environment A* Agent Path.
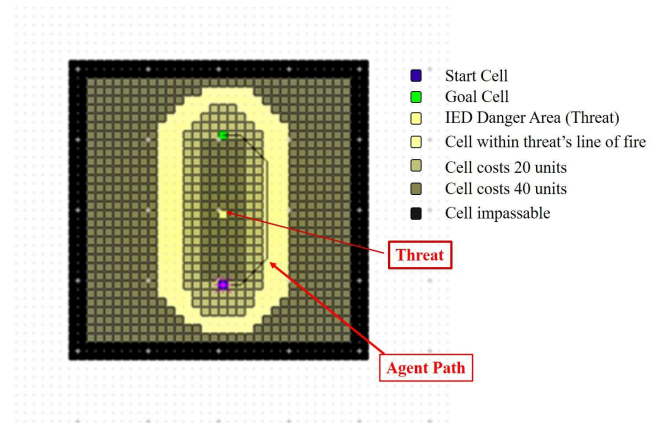Source: Matthews and van der Sterren (2002).

Figure 7.   Hot Temperature Environment A* Agent Path.
Source: Matthews and van der Sterren (2002).

B.   MODEL IMPLEMENTATION

The process for development began with building the desired scenario in a small modeling environment covering a 100 by 100-meter area of terrain with a 100 x 100-grid

square waypoint graph. Each square cell is sized at two by two meters with the basic capability of demonstrating similar behavior as in the *A* Explorer* implementation. Following successful completion of this build, the next build involved expanding the model's waypoint graph to cover a 500 by 500-meter terrain space in order to provide more realistic offset distances for the agent's navigational path in accordance with the recommended offset distance. This waypoint graph consisted of a 100 x 100 grid with cell sizes of ten by ten meters.[17] The final stages of the model build entailed implementing random variability within the A* algorithm to ensure variable-sized offset from the IED danger area in the paths of the agent during each trial execution of the model, as well as a version capable of executing experiments with a massive number of trial runs (10,000 trials for each temperature condition). The *Apex Path* pathfinding library served as the mechanism for implementing the waypoint grid and A* algorithm, as well as provided a model for the agent (Apex Game Tools, 2016; Apex Game Tools, n.d.). The source for modeling the danger area consisted of models provided by the *Modular Prison Fortress* library (Aquarius Max, 2016).

Inducing variability into the agent's generated navigational paths provided a mechanism for implementing the statistical findings related to human vigilance and attention behavior in hot environments in such a way that proved realistic. Due to the vast number of possible influences on a human's behavior, allowing for variability in paths such that even under mild temperature conditions an agent could walk paths that moved dangerously close to a potential IED danger area becomes necessary to ensure realism. Likewise, an agent in a hot temperature environment could also find a navigational path in accordance with recommended practices and navigate from the recommended standoff distance away from the IED danger area at certain points in time. The inclusion of a mechanism for variability was essential for ensuring that the proposed model proof-of-concept provides instances where agents make both good and bad decisions, but more

---

[17] The object forming the navigational terrain consisted of a 0.5-meter side length cube scaled up by a factor of 1000 in the x and z dimensions to achieve a 500 by 500-meter terrain space. The scale in the y dimension was one. In the small model environment, this 0.5-meter side length cube was scaled up by a factor of 200 in the x and z dimension to achieve a 100- by 100-meter terrain space.

often than not make decisions that replicate performance described in Hancock et al.'s (2007) meta-analysis.

1. A* Algorithm Modifications

Prior to implementing variability within the model based upon Hancock et al.'s (2007) thermal stressors meta-analysis, the underlying implementation of the A* algorithm required manipulation to promote navigation close to the danger area under hot temperature conditions, and far from the danger area under mild temperature conditions. The mechanism for manipulating the A* algorithm required adding an additional cost to each node within the waypoint graph via the "g" portion of the node's total "f" cost. Doing so within the "g" portion of the cost, as opposed to the "h" portion of the cost, allowed for more responsive alteration of the agent's path. After initial attempts to fine tune the cost functionality within the "h" portions of the cost as well, evidence suggested that this was unnecessary and at times did not generate the desired decremented path finding behavior. Modifications to the "g" portion of the cost adheres to techniques described by Stout (2000), and demonstrated by Darken (2016) as well as Matthews and van der Sterren (2002). Furthermore, using a heuristic "h" cost based upon the "straight-line distance" between the current node and the goal node (as we do in our model) allows us to preserve A* graph search optimality according to Russell and Norvig (2010), who stated that such a straight line heuristic "is a consistent heuristic" (p. 95).

a) Model Code Structure and Implementation

The code implementation for the A* algorithm took place within the *MyMoveCostFactory* component of the *Game World* object. The *Game World* serves as the *Apex Path* mechanism for providing a waypoint grid, as well as the object basis for adding additional components for executing the A* algorithm and visualizing the resultant agent paths (Apex Game Tools, 2016). *MyMoveCostFactory* and the underlying *MyMoveCost* component held the main code for altering the agent's "g" portion of the cost, and serve as the components within *Apex Path* for tailoring the A* algorithm to one's own desired cost influencers (Apex Game Tools, 2016). The Modeling, Virtual Environments, and Simulation (MOVES) support staff provided some assistance in the

form of code that included the initial desired concept for agent paths according to the *A* Explorer* implementation within *MyMoveCostFactory* and *MyMoveCost*—we built upon this for our model implementation (Apex Game Tools, 2016).

The additional enhancements and modifications to *MyMoveCost* created by the MOVES support staff allowed for adjustments to the agent's navigational paths to include variable lateral offsets from the IED danger area primarily through the *GetEnvironmentCost* function. Essentially, after the randomly generated establishment of a circular danger area radius value (discussed in Section 1b) Assigning Values to *randomNumberDASize*), called the *randomNumberDASize* (corresponding to the IED's circular danger area radius), the *GetEnvironmentCost* function calculates the "g" cost of the currently searched node based upon the node's distance relationship from the IED danger area. In most cases, if the node's location fell within the circular danger area radius the "g" cost would receive a minimum additional cost of 10,000 units, discouraging the A* algorithm from choosing a path that would pass through the IED's circular danger area radius. However, if the node's location fell outside of the IED danger area radius, the "g" cost would increase based upon slightly different methods of calculation due to the environmental temperature condition. In a hot environment, the "g" cost penalty increases from the edge of the circular IED danger area radius outward to the edge of the waypoint grid, while for the mild temperature environment the "g" cost penalty decreases from the edge of the IED circular danger area radius outward to the edge of the waypoint grid.

Figure 8 depicts the flow of information within the model to achieve the desired temperature decremented path finding behavior. The execution of the model begins with a user set Boolean variable to define the environment's status as a hot or mild temperature environment. This variable then drives code execution within the *EnvironmentStatus* component, and results in different paths based upon random number draws from two different normal distributions. If the agent operates within a hot temperature condition environment, the model utilizes a pseudorandom number generator to sample from a normal distribution describing *performance accuracy* in a hot temperature environment. However, if the agent operates in a mild temperature condition

environment, the model utilizes a pseudorandom number generator to sample from a different normal distribution related to *performance accuracy* in a mild temperature environment (Section 2a)   Derivation of Hot and Mild Temperature Condition Distributions) describes the derivation of these two normal distributions). The random number draw sample is then compared to where it falls into on the normal distribution for the mild temperature condition to assign a value to the *randomNumberDASize* variable. The assigned value of *randomNumberDASize* within the *MyMoveCost* script informs the *MyMoveCostFactory* component the correct offset value to utilize in assisting with the calculation of a searched node's "g" cost (Apex Game Tools, 2016). The *PathFinderVisualizer* component assists in executing the A* algorithm for this model implementation and displays the ultimate found path; it also depicts the total "f" cost of each searched node via a heat map visualization (Apex Game Tools, 2016).
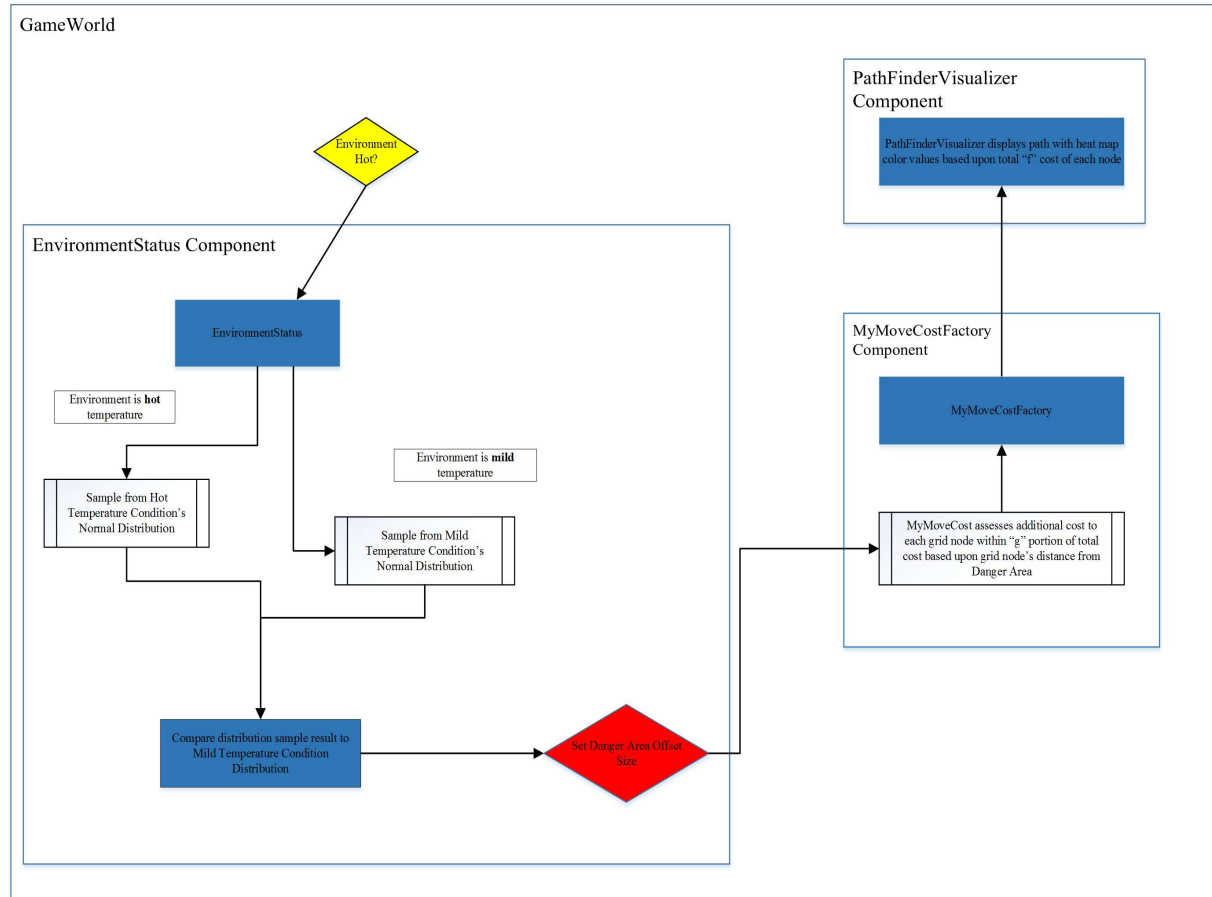
Figure 8. Flow Diagram for Depicting Decremented Pathfinding Behavior

b) *Assigning Values to randomNumberDASize*

We split the mild temperature condition's normal distribution into "brackets" by splitting the distribution up into half-standard deviation increments between negative three standard deviations from the mean to positive three standard deviations from the mean. Different values get assigned to the *randomNumberDASize* variable based upon the location of the randomly generated number from the respective environmental condition normal distribution as compared to the bracket location on the mild temperature condition's normal distribution that the sample would be found. Figure 9 and Figure 10 show the two normal distributions utilized for the model implementation. The brackets can be seen in Figure 10. The text box below the x-axis displays the possible values assigned to the *randomNumberDASize* variable. No interpolation of *randomNumberDASize* takes place between brackets.

Smaller danger area values exist at the left end of the distribution, coinciding with poorer performance accuracy indicating poor attention and vigilance, and thus less offset achieved by the agent from the IED danger area. Larger danger area values on the right side of the distribution indicate better performance accuracy due to better attention and vigilance in the respective temperature environment, and consequently more offset of the agent's path from the IED danger area. This implementation provides a means for linking human statistical performance as seen in Hancock et al.'s (2007) meta-analysis to tactical path-finding decisions by the agent.
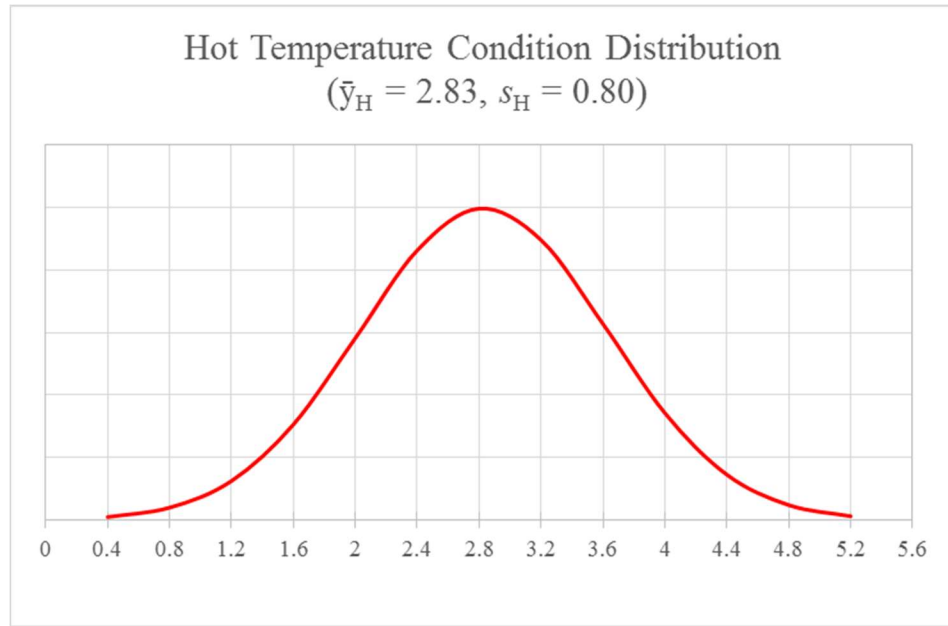
Figure 9.   Hot Temperature Condition Distribution for Performance Accuracy.
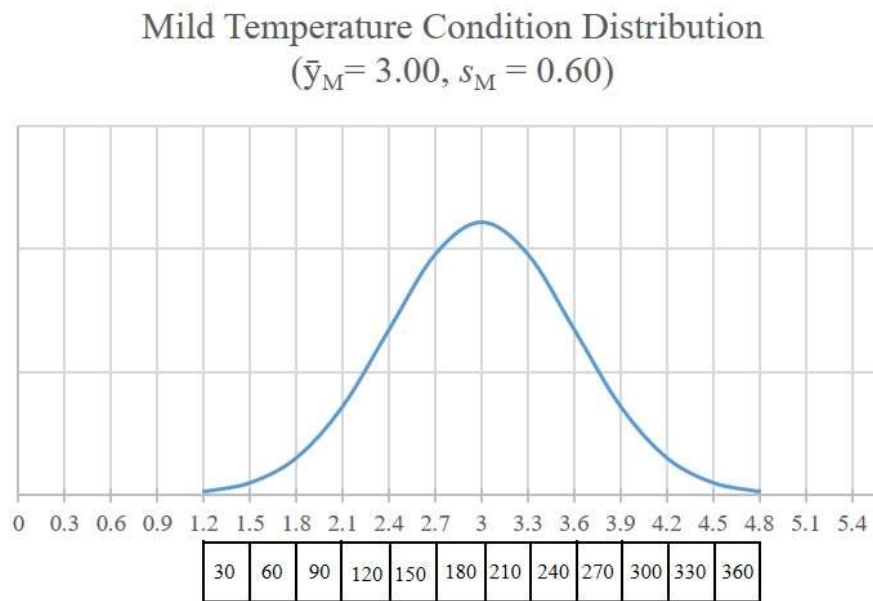Adapted from Hancock et al. (2007).

Figure 10.   Mild Temperature Condition Distribution for Performance
Accuracy. Adapted from Hancock et al. (2007).

The circular danger area radius generates the most notable feature of the agent's path—the path's lateral offset from the IED danger area at the path's midpoint. The influence of the different "g" cost calculation method's due to the environmental temperature condition results in differences to both the A* algorithm's search space size and aesthetic differences in path appearance. The goal for calculating the additional "g" cost in different ways for each of the temperature conditions centers on the idea of promoting paths more offset from the danger area under mild temperature conditions, and less offset from the danger area under hot temperature conditions. This approach proved to have a smaller impact on ultimate path generation than originally thought. In execution, a preponderance of the determination of the final path appearance rests with the determination of the circular IED danger area radius.

We include a number of examples of the agent paths generated under the two temperature conditions. Figure 11 shows the initial model environment set up in Unity. The left side of the model environment (denoted by "Scene View") shows the initial model set up from an object standpoint, while the right view (denoted by "Game View") shows the view seen by a user in a gaming environment. The small light blue object denotes the agent; while light blue objects appear in both views, there is only one object—each viewport is giving a different view of the same object. The colored lines on the terrain on each side of the IED danger area location indicate the various *randomNumberDASize* offset "test gates" for assistance with model testing. The "Environment Hot" check box toggle in the top left of the game view allows the user to "alter" the environmental temperature of the model via the underlying statistical implementation of Hancock et al.'s (2007) meta-analysis.

Figure 12 and Figure 14 show sample path outputs for an agent in both hot and mild temperature condition settings. Figure 12 displays an agent's path in a hot temperature environment with an assigned value for *randomNumberDASize* of 150 meters. The console view (denoted by "Console View") displays the distribution sample (called *distSample*) taken via random number generator from the hot temperature condition normal distribution; since *distSample* fell into the -0.5 to -1.0 standard

59

deviation bracket on the mild temperature condition normal distribution, *randomNumberDASize* receives a value of 150 meters. The agent's path is depicted in blue, while the A* search space is the dark blue area outlined in bright green. The bright green indicates nodes awaiting processing by the A* algorithm (Apex Game Tools, 2016). The circular gray area in the middle of the search space indicates the 150-meter circular area described by *randomNumberDASize*; these nodes are not searched by the algorithm likely due to their extremely high total "f" cost relative to the dark blue nodes that do get searched.

In Section 1d) Heat Map Visualization of Search Space, we will discuss the assignment of colors to the searched nodes, as each color represents different "f" costs for each node. Figure 14 displays similar model functionality to that shown in Figure 12, but in a mild temperature environment. We include Figure 13 and Figure 15 with the search space color in light gray to provide better visibility of the path behavior seen in Figure 12 and Figure 14.
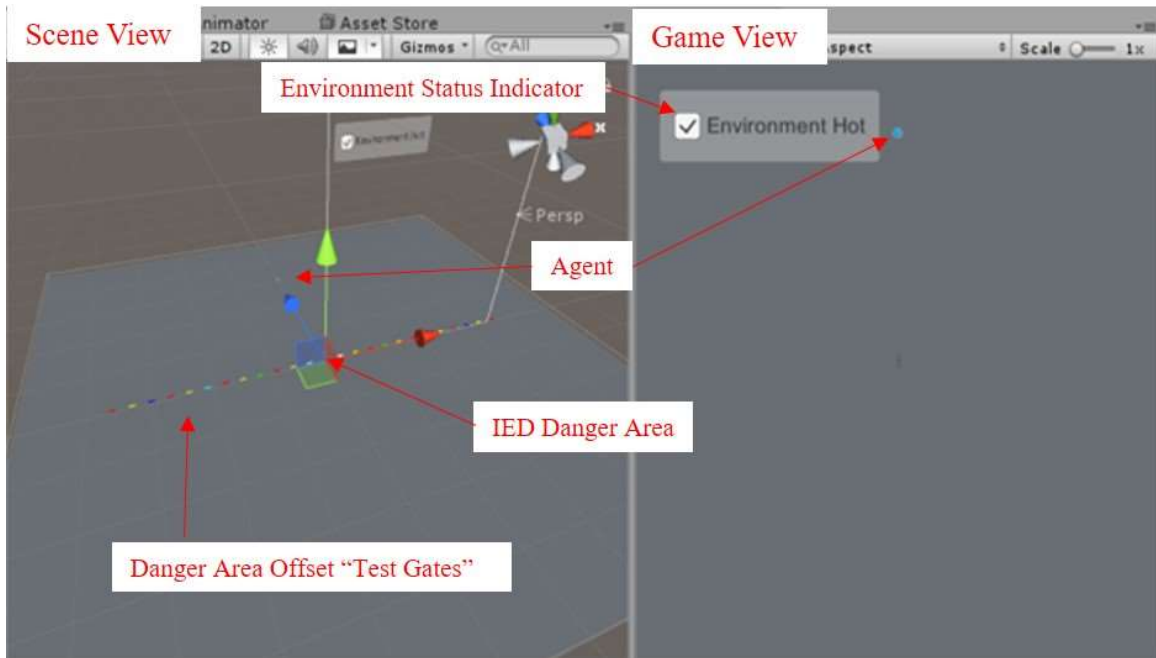

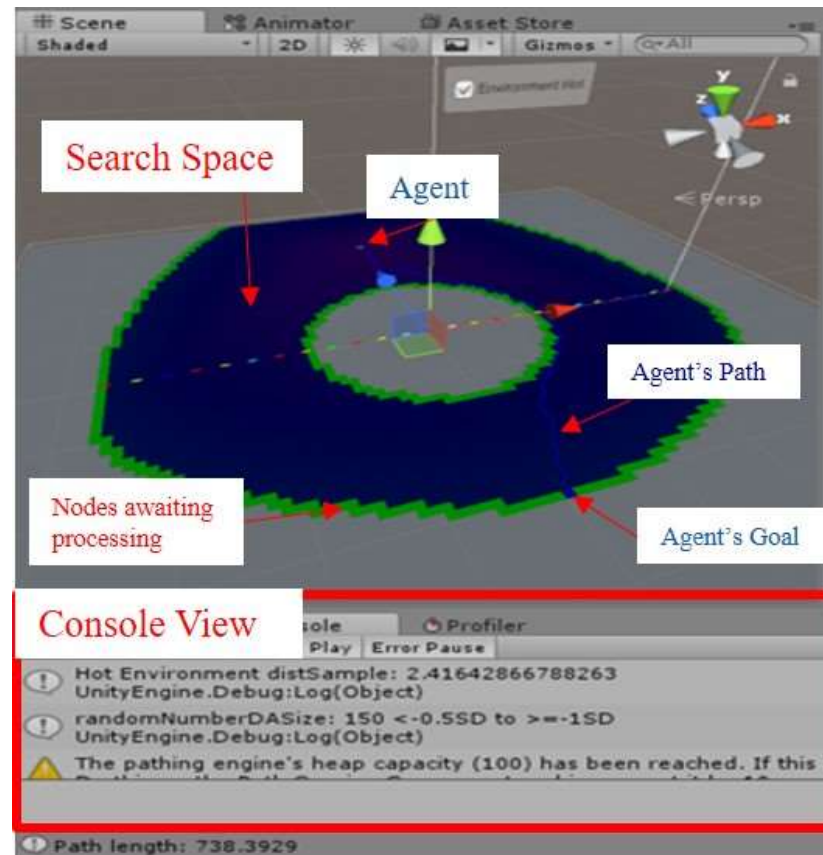Figure 11.   Initial Model Set-up in Unity

Figure 12.   Model under Hot Environmental Conditions: Agent Selects Path with 150 Meters Offset from the IED Danger Area.
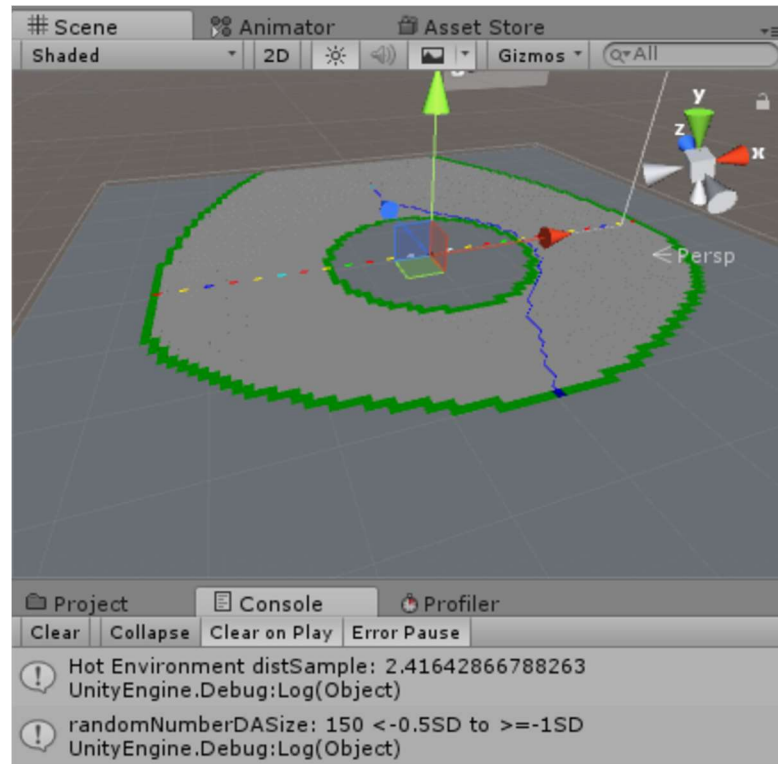
Figure 13.   Hot Environmental Conditions: 150 Meter Offset Blue Path on Light Gray Search Space
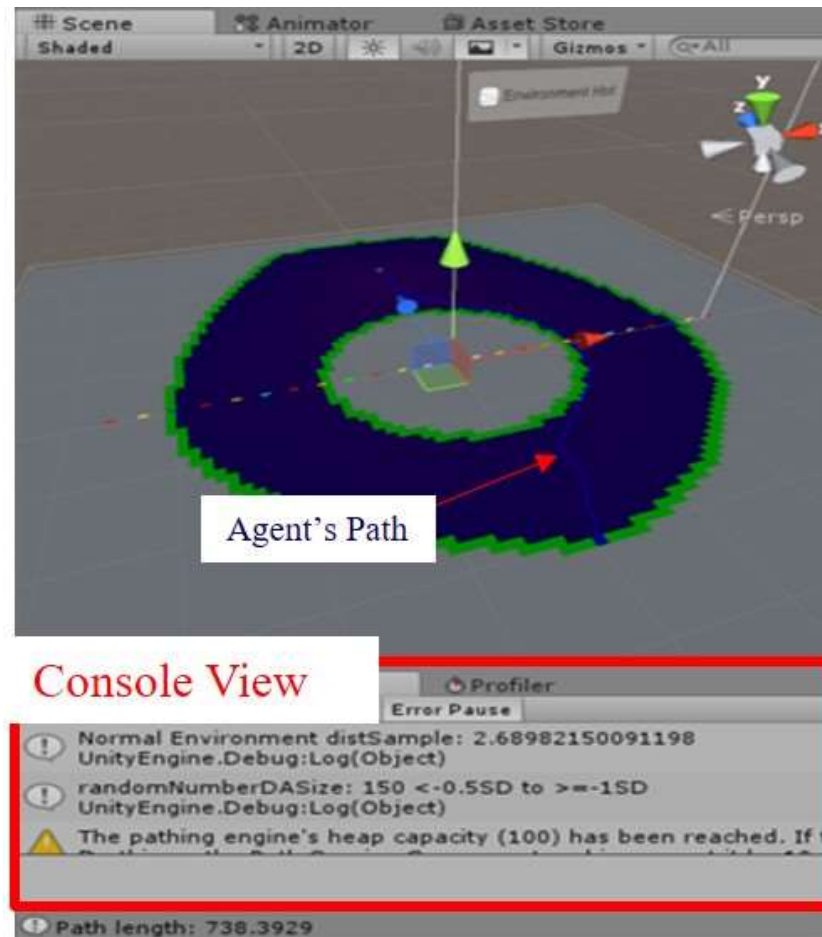
Figure 14.   Model Under Mild Environmental Temperature Conditions: Agent Selects Path with 150 Meters Offset from the IED Danger Area.
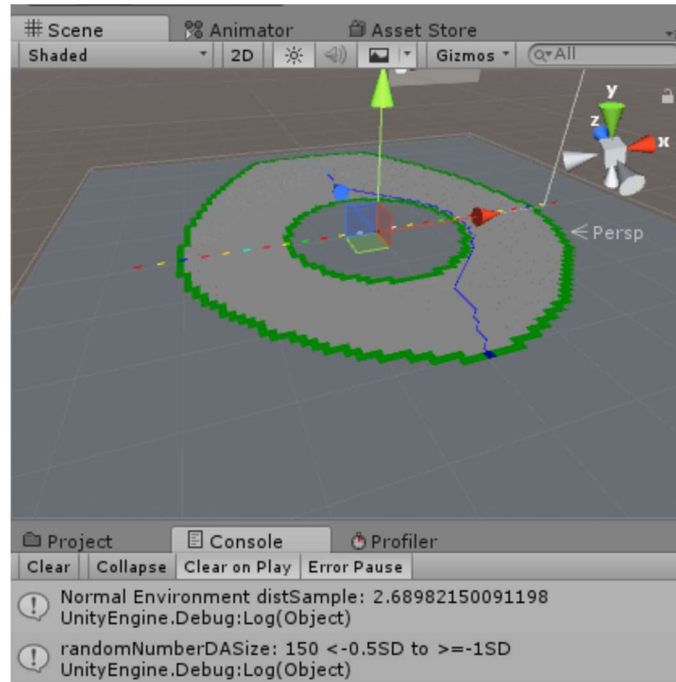
Figure 15.   Mild Environmental Conditions: 150 Meter Offset Blue Path on Light Gray Search Space

As depicted in Figure 13 and Figure 15, the widest point of both agent paths corresponded to the selected *randomNumberDASize* value resulting from the model's algorithm for implementing randomness based upon the relation of a sample from the respective temperature condition's normal distribution to the mild temperature condition's distribution. In the cases shown in Figure 13 and Figure 15, *randomNumberDASize* received a value of 150, meaning that all nodes within a radius of 150 meters of the IED danger area received such a high "g" cost that the A* algorithm found other nodes to search that proved much cheaper. One can also see how the path generated in the hot temperature condition appears visually different from the path generated in the mild condition (specifically after the agent's path has moved beyond the circular IED danger area radius). This results from the slightly different calculations of the "g" cost of nodes outside of the unsearched dark gray circular space around the IED danger area. Additionally, the size of each condition's search space constitutes another noticeable difference between the two conditions, where the mild temperature condition exhibited a smaller search space for this IED danger area offset size radius of 150 meters.

d) *Heat Map Visualization of Search Space*

The *Apex Path* tool provides a mechanism via the *PathFinderVisualizer* component to view the various aspects of each searched node's cost (Apex Game Tools, 2016). Figure 16 displays the "f" cost for each node searched for the agent's path found under the same hot environmental conditions and IED danger area offset size as shown in Figure 12. The default color coding implementation of the *PathFinderVisualizer* component for search space nodes received alteration by using a heat map color assignment function (Darken, 2016) to assign color values from the Hue-Saturation-Value (HSV) color wheel (skm, 2014) to each searched node in the search space based upon its total "f" cost. Figure 16 shows a close up view of the assigned "f" costs for a section of searched nodes. The game view displays the total "f" cost for each searched node within a small section of the search space covering the edge of the IED danger area radius and the search space area through which the agent's path travels.
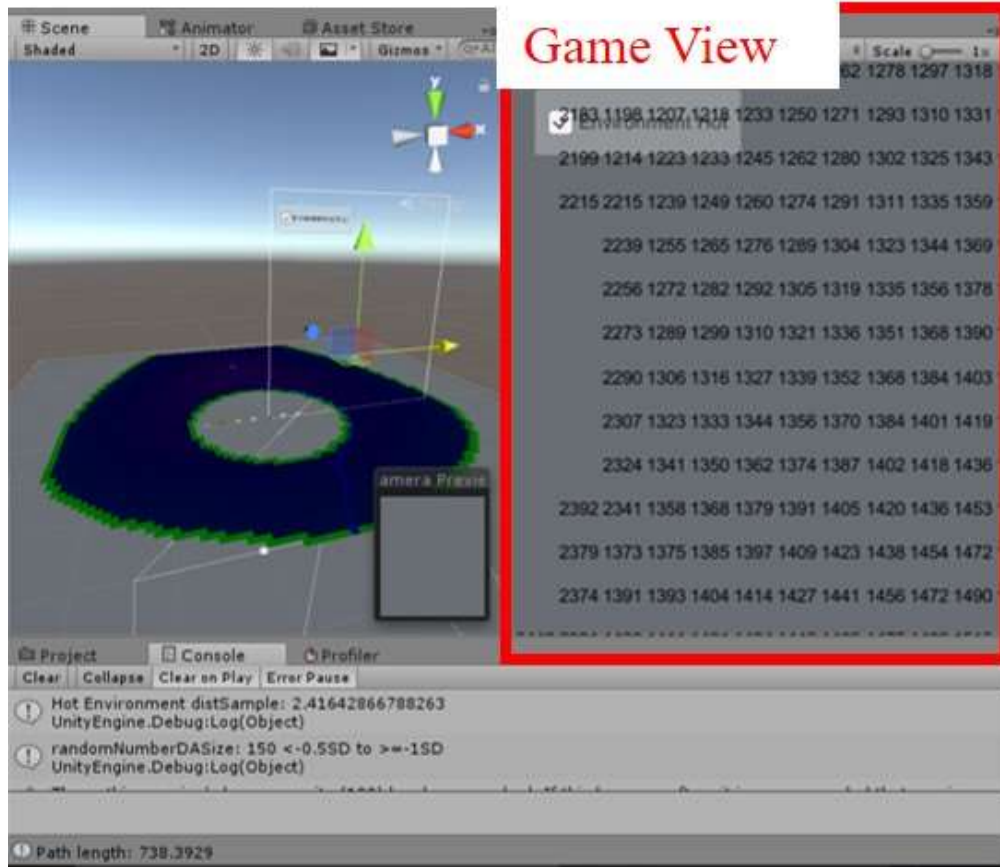
Figure 16.   Game View of Hot Environment Condition Displaying Total "f"
Cost for Each Searched Waypoint Node

The heat map generally displays dark blue hue values for many of the agent paths for a smaller circular IED danger area (240 meters and smaller); indicating low overall cost for the search nodes. However, as the circular IED danger area radius grows to 270 meters or larger, the heat map values begin to display warmer (brighter green and red) color values indicating more varied node costs. Figure 17 through Figure 24 (except for Figure 18 and Figure 20, which we include to make it easier to see 270 meter offset paths than those shown in Figure 17 and Figure 19) show this change in color values.

An agent's chosen path has a lateral offset from the danger area similar to the value chosen for the circular IED danger area radius. However, the paths generated in environments where the *randomNumberDASize* value received a value of 330 or 360

differ in that the underlying search space indicates that multiple nodes inside of the chosen IED danger area radius get searched (see Figure 21 through Figure 24).

The presence of more color variation and change in the hot temperature conditions with 270 meter and larger IED danger area radii results from the different implementations for calculating the additional "g" cost between the hot and mild temperature conditions. While the additional "g" cost increased in size from the danger area outward to the *Game World's* fringes for the hot temperature condition, in the mild temperature condition the additional "g" cost decreased from higher values in vicinity of the danger area to lower values at the edge of the *Game World*. Additionally, the added "g" cost increased in smaller and more consistent increments in the hot temperature condition, i.e., the cost value increased one unit for every ten additional meters of distance between the node and danger area, with interpolation of cost increases in between. The mild temperature condition additional "g" cost decreased in less consistent steps than the hot temperature condition did—for example, the added "g" cost at nodes 300, 350, 400, 450, and 500 meters, respectively, from the IED danger area were 100, 90, 70, 40, and 22.5. This differing calculation likely contributes to the bigger search spaces seen in the hot temperature condition paths, as well as the wider range of heat map colors (indicating a wider range of "f" costs) seen outside of the circular IED danger area radius corresponding to the value chosen for *randomNumberDASize*.

As an example, in the search space for the hot temperature condition path shown in Figure 21, heat map values appear reddest where the "f" cost is highest around the fringes of the circular area of gray unsearched nodes. The green colors indicate nodes whose "f" costs are moderately high, i.e., higher than the low "f" cost blue nodes around the agent's start point, but lower than the high cost red and yellow nodes. One can also see how the search space for the mild temperature condition path shown in Figure 22 displays fewer red and yellow nodes, and generally has more green and blue nodes than the search space for the hot condition shown in Figure 21. In the search spaces in Figure 23 and Figure 24, some nodes around the circular area of gray unsearched nodes appear to form a black ring. This black ring corresponds to nodes with high enough "f" costs that the heat map color function failed to map the node to a value on the HSV color wheel.

One can see how the search space in the hot temperature condition shown in Figure 23 displays a wider black ring of nodes and many more red color values than the search space from the mild temperature condition shown Figure 24, indicating much higher and varied costs in the hot temperature condition compared to the mild temperature condition.
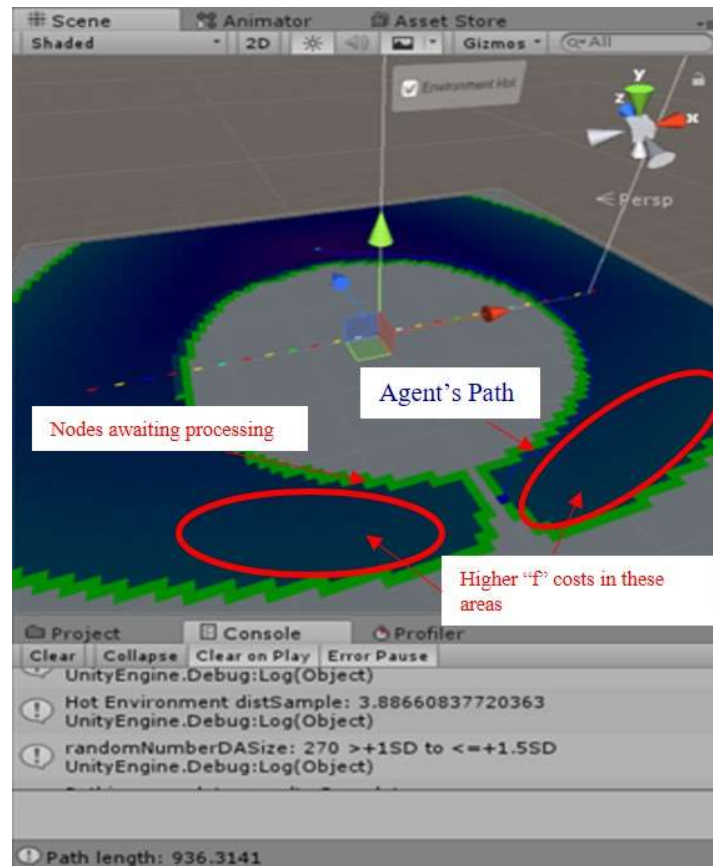
Figure 17.   Heat Map Display for Hot Temperature Condition 270 Meter Radius Path

Figure 18.   Hot Temperature Condition: 270 Meter Offset Blue Path on Light Gray Search Space
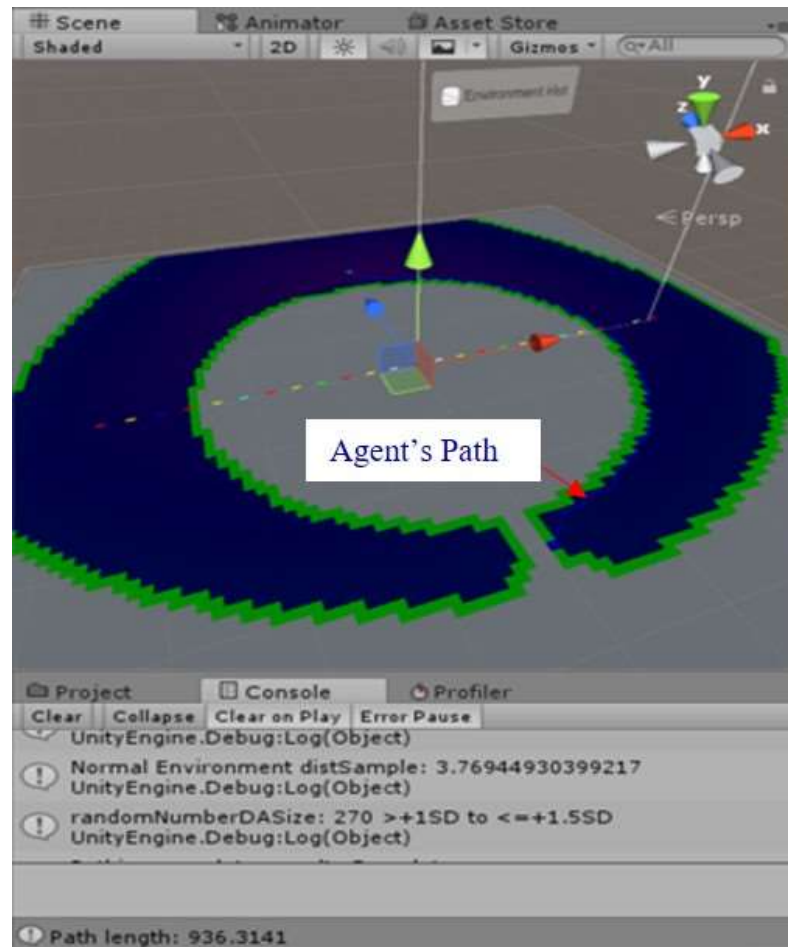
Figure 19.   Heat Map Display for Mild Temperature Condition 270 Meter Radius Path

Figure 20.   Mild Temperature Condition: 270 Meter Offset Blue Path on Light Gray Search Space

Figure 21. Heat Map Display for Hot Temperature Condition 330 Meter Radius Path

Figure 22.   Heat Map Display for Mild Temperature Condition 330 Meter Radius Path
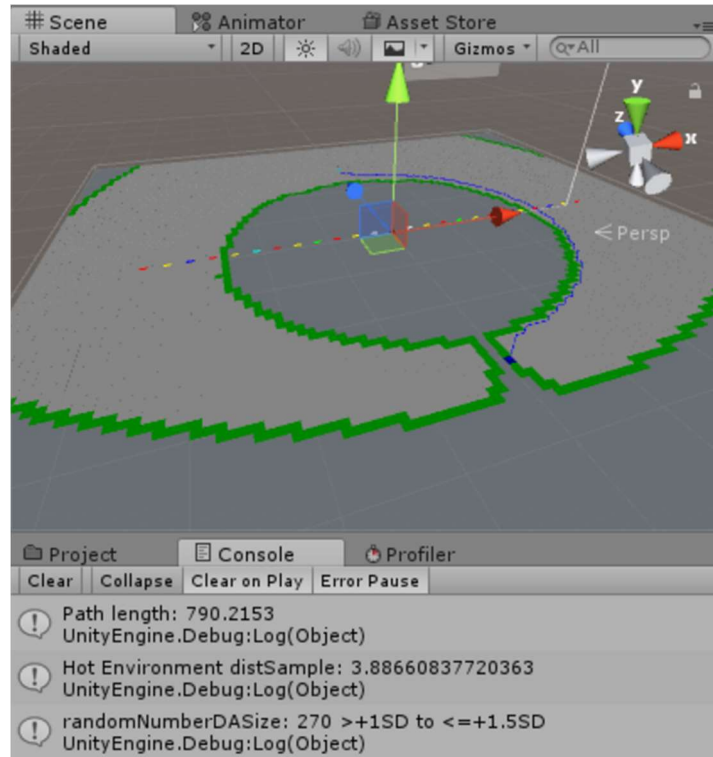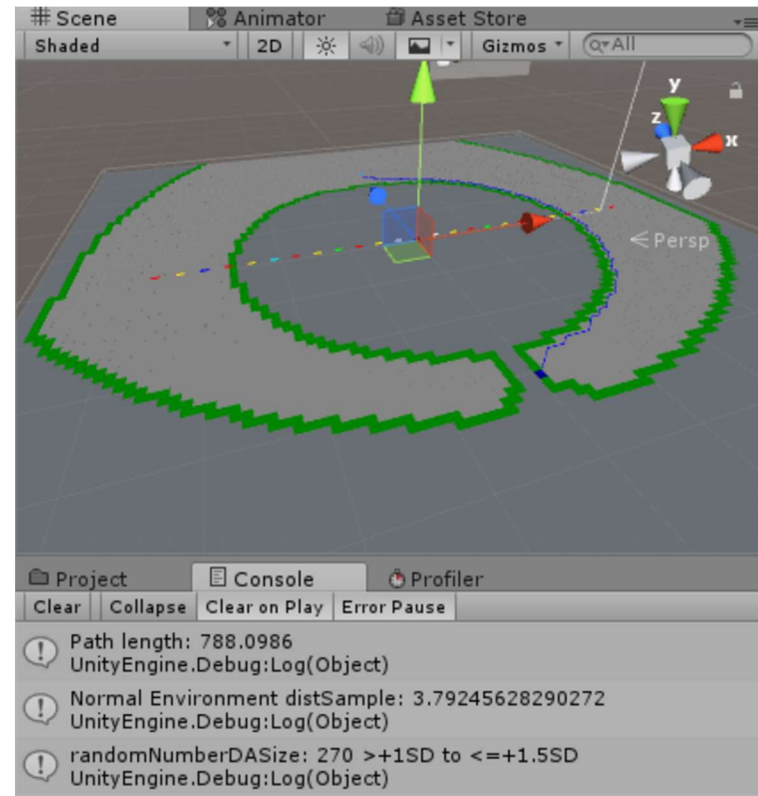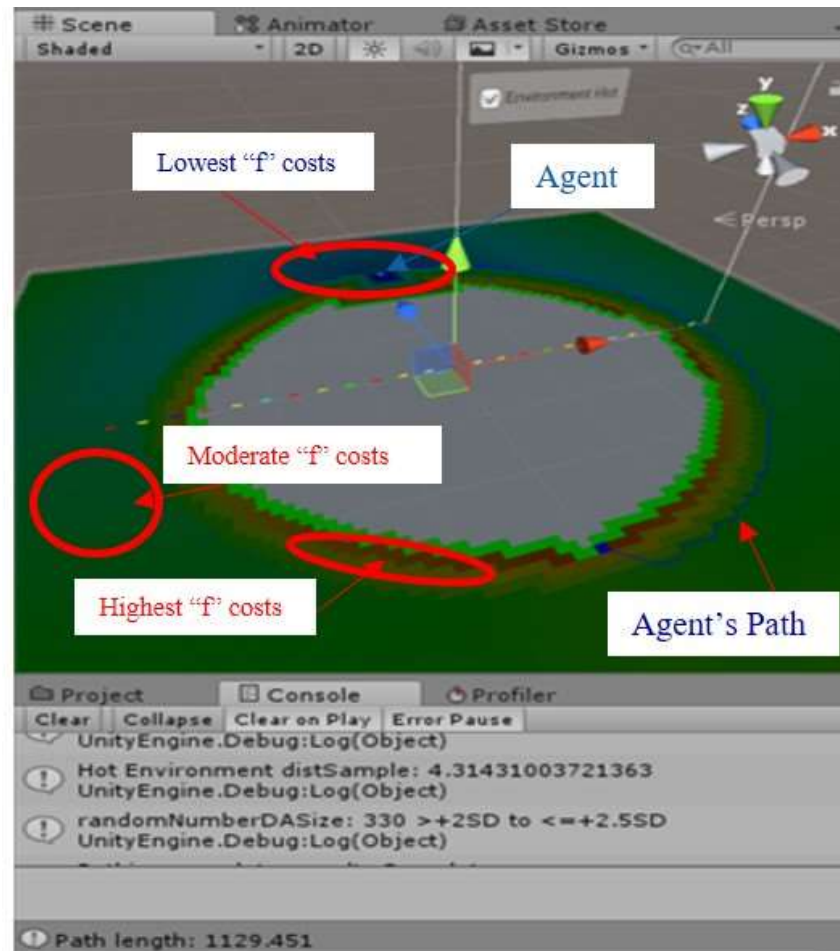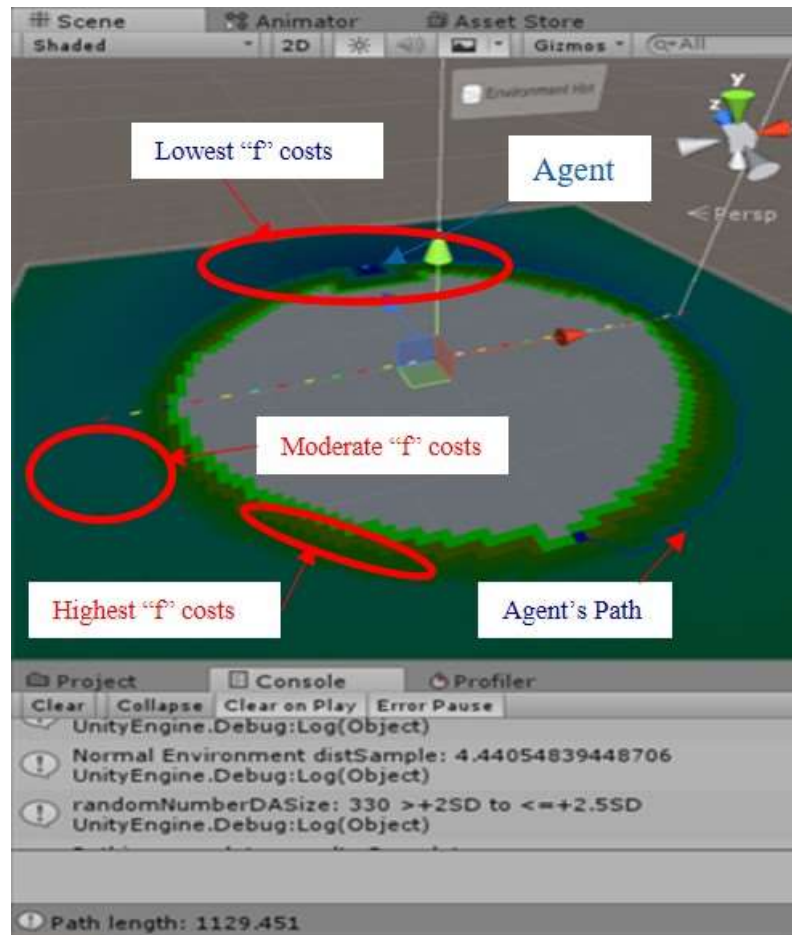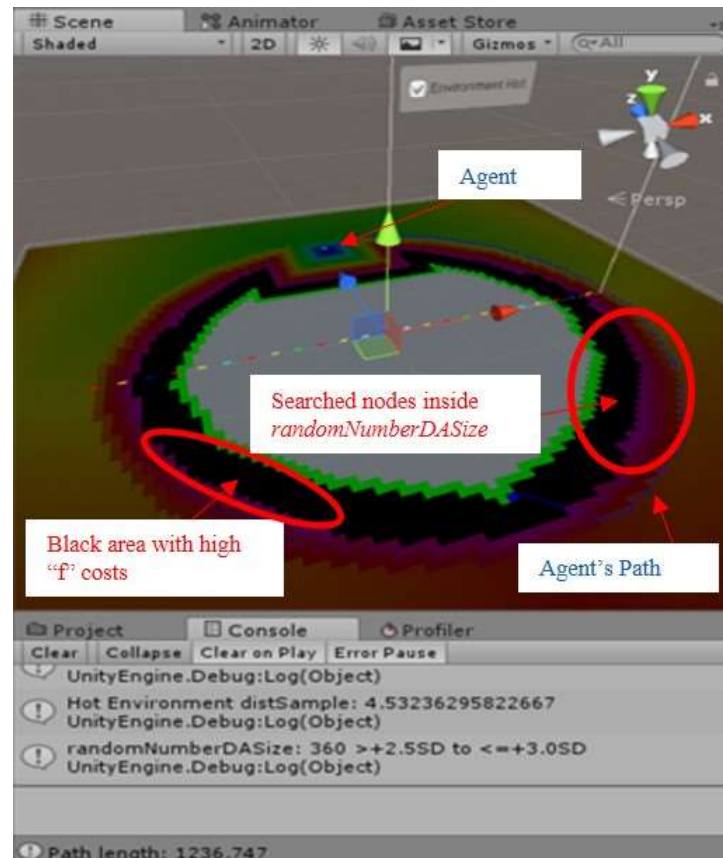
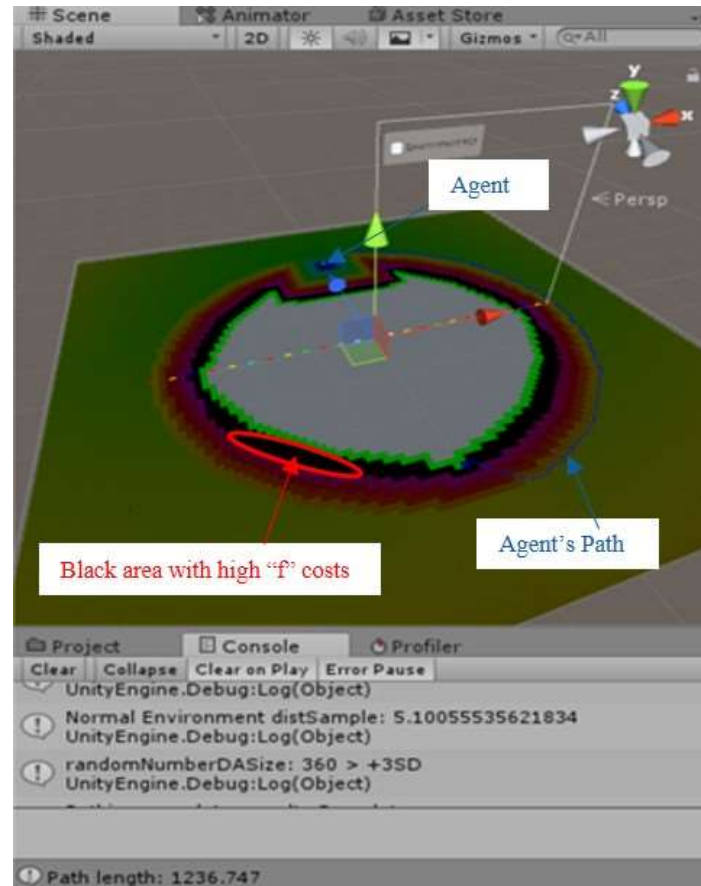Figure 23.   Heat Map Display for Hot Temperature Condition 360 Meter Radius Path

Figure 24.   Heat Map Display for Mild Temperature Condition 360 Meter Radius Path

2. Design of Experiments

We now provide a discussion of the experimental design for this thesis for the purposes of describing how we tested for differences in paths between the hot and mild temperature conditions.

a) Derivation of Hot and Mild Temperature Condition Distributions

In order to create an empirically based behavior model, we used Hancock et al.'s (2007) meta-analysis to provide the foundation for creating normal distributions for each temperature condition to sample from and achieve variability in the path output. To create the normal distributions (Figure 9 and Figure 10), we used the statistical mean effect size of negative 0.29 from the outlier analysis of the perceptual task category conducted in greater than 85°F ET as the basis for an equation to calculate the appropriate sample means and standard deviations for the respective temperature condition normal distribution (Hancock et al., 2007, p. 858). The negative value represents worse performance for the "experimental" hot temperature condition relative to the "control" mild temperature condition (Hancock et al., 2007, p. 855). The equation utilized to calculate the means for the experimental and control conditions was:

$$ g = \frac{\bar{y}_H - \bar{y}_M}{s} \tag{1.2} $$

Hancock et al. (2007) utilized a form of this equation to calculate the effect size for specific studies, where g represents a specific study's effect size, s represents the study's standard deviation, $\bar{y}_H$ represents the mean value for the hot temperature condition, and $\bar{y}_M$ represents the mean value for the mild (i.e., "room-temperature") temperature condition (p. 854). Hancock et al.'s (2007) methods assume s represents a pooled standard deviation for studies utilizing a "between-subjects" experimental design, and a standard deviation of the control condition for a "within-subjects" experimental

design (p. 855).[18] Since the number utilized for the effect size of negative 0.29 represents a mean weighted effect size that required application of multiple statistical techniques by Hancock et al. (2007), specific standard deviations did not appear in the meta-analysis. Furthermore, the researchers "adjusted" each standard deviation to a "common standard deviation…of the experimental and control groups" before calculating the mean weighted effect size, so for the purposes of this thesis, we assume that s represents the mild temperature condition's standard deviation (p. 855). We selected 0.6 for s to allow for a randomly generated sample small enough that sampling the normal distribution at greater than plus or minus three standard deviations from the mean only rarely produced a negative value.

Due to a lack of information present in the meta-analysis to reverse the process used to calculate the mean weighted effect size, we assumed that the mean weighted effect size of -0.29 can be substituted for g in Equation 1.2. Finally, by solving Equation 1.2 for the numerator on the equation's right side we are left with:

$$\bar{y}_H - \bar{y}_M = -.174 \tag{1.3}$$

We then created mean values representing mean performance accuracy for the hot and mild temperature conditions that would fulfill this equation, ending up with $\bar{y}_H = 2.83$ and $\bar{y}_M = 3.00$. To calculate the standard deviation for the hot temperature condition, we considered that "performance under thermal stressors proved on average to be approximately one third of a standard deviation…worse than performance at a comparative thermoneutral temperature" (Hancock et al., 2007, p. 857).[19] Therefore, we chose 0.8 as the standard deviation for the hot temperature condition. The desire to calculate the means and standard deviations as representative of performance accuracy stems from Hancock et al.'s (2007) finding that "heat stressors deleteriously influence

---

[18] Hancock et al. (2007) applied other statistical methods to create effects sizes if they did not have access to the "means and/or standard deviations" (p. 855).

[19] A limitation of this finding is that "thermal stressors" includes both hot and cold stressors, while our model only accounts for hot stressors (Hancock et al., 2007).

perception through a reduction in response accuracy and an increase in response time" (p. 868). The authors also note from the outlier analysis a "high degree of consistency in the effect of heat on perceptual performance accuracy" (p. 869–870).

b) *Approach for Creating Decremented Behavior in Hot Environment*

The normal curves for the hot and mild conditions Figure 9 and Figure 10 show the splitting of the distributions at every half of a standard deviation and the mappings of those increments to a value for the IED danger area offset radius. The decision to account for values randomly sampled from each temperature condition's distribution up to plus or minus three standard deviations from the mean stems from the "68-95-99.7 Rule**,**" where plus or minus three standard deviations contain 99.7% of the normal distribution's values (De Veaux, Velleman, & Bock, 2012, p. 130). Any randomly sampled values that fell outside of the range were treated as if they had fallen within the range of the nearest two-and-a-half to three standard deviations from the mean bracket (Oferei, 2014).

In order to create the desired behavior of decremented performance in a hot temperature environment that aligned with the extant human performance research, it became necessary to design the model to compare the random number generated samples from both temperature conditions to the mild temperature condition serving as the control condition. We utilized the *MathNetForUnity* package to provide the Mersenne Twister random number generator, as well as the mechanism for conducting random number sampling from normal distributions (Bismur Studios Ltd., 2017a; Bismur Studios Ltd., 2017b; Matsumoto & Nishimura, 1998). The *MathNetForUnity* package utilized the Box-Muller Algorithm (Box & Muller, 1958). We created a model version for executing mass trial runs for statistical testing. Statistical results for path length (in meters), the random number generator sample from the respective temperature condition, and the value chosen for *randomNumberDASize* were collected via the *ExperimentStats* Unity Package created by the MOVES support staff.

c) *Controlling for Variability*

Several approaches were used to control for variability. To reduce variability while measuring the length of time needed to execute trial runs, we only allowed the

Unity model and Visual Studio project window to remain open. Further, the trials for each temperature condition took place during separate runs of the model. This resulted partly from the structure and implementation of the code, and partly for ensuring that executions of the model in one temperature condition did not affect executions of the model in the other temperature condition, or negatively affect the time to execute runs.

Another approach utilized to control for variability came from the use of the "common random numbers (CRN)" technique described by Law (2015) (p. 588). Essentially, we used the same seed for the Mersenne Twister pseudo-random number generator to sample from each temperature condition's normal distribution. According to Law (2015), in doing so we hope to "induce…positive correlation" (p. 589) between each temperature condition's outputs, ultimately yielding less variance and a "much smaller confidence interval" of the mean of the differences between the two conditions (p. 563). Law (2015) describes how using CRN mandates the use of paired-t tests of the outputs between the two temperature conditions due to a lack of independence between each temperature condition's corresponding outputs. Further, we conducted the same number of trials for each temperature condition—another indication that we needed to use the paired-t test to analyze the model's results (Law, 2015). The CRN technique removes the variance from the model arising from having different numbers seeding the Mersenne Twister used in each temperature condition; the author states that the technique essentially allows one to "compare the alternative" temperature conditions "under similar experimental conditions" (p. 589). Additionally, the author relates that this approach does not alter the "means and variances" of the normal distributions of each temperature condition (p. 589).

d)   *WBGT Flag Condition Indicator*

To aid in comprehension of the environmental temperature conditions in a manner familiar to Marines, we placed additional features to describe the environment in terms of environmental flag conditions above the agent. The Marine Corps utilizes a system of environmental flag conditions to describe the current WBGT of the environment; this system provides "guidance" for outdoor training in hot environments depending on the

environmental flag condition (Commandant of the Marine Corps, 2003, p. 5). Since 85°F ET corresponds to 87.4° WBGT (Hancock et al., 2007), the hot temperature condition displays a red flag condition sphere above the agent to depict a hot environment, while a white flag condition sphere displays above the agent in a mild temperature environment. Although the 85°F ET/87.4°F WBGT threshold should technically fall into the yellow flag condition WBGT range (85°F-87.9°F WBGT), we assume that it falls into the next highest flag condition range of red (88°F-89.9°F WBGT) (Commandant of the Marine Corps, 2003). The white flag condition technically encompasses temperatures up to (but not including) 80°F WBGT, and although this thesis equates the mild temperature condition to room temperature per Hancock et al. (2007), we use white here to describe the mild temperature condition (Marines.mil, n.d.). Figure 25 and Figure 26 depict the visualization features used to describe the environmental temperature condition of the agent's operational environment.
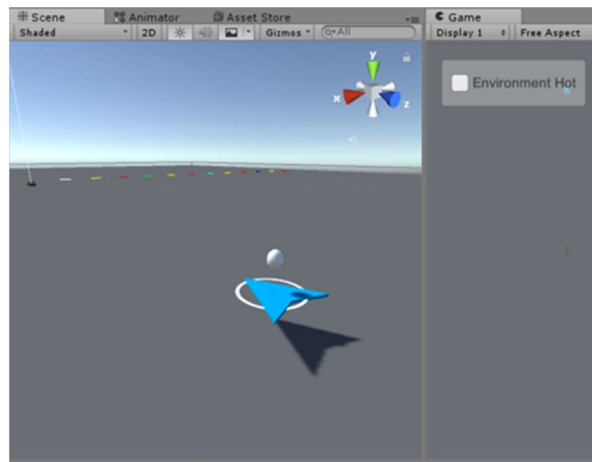

Figure 25.   White Sphere above Agent Indicating a Mild Temperature Condition Operational Environment
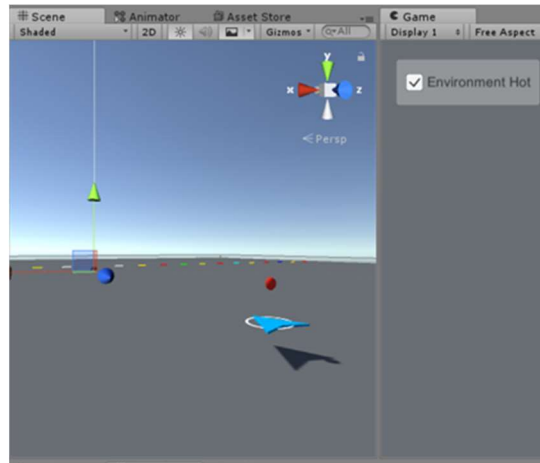
Figure 26.　Red Sphere above Agent Indicating a Hot Temperature Condition
Operating Environment

IV.   RESULTS

A.   ASSUMPTIONS AND CONDITIONS

We will now discuss the hypotheses utilized to help answer one of our initial research questions as to the extent to which we could quantify human performance in an AI model within a military training simulation. We present our findings from the testing of each hypothesis for runs of the model under each temperature condition. The findings presented will include results for testing of the hypothesis with an experiment with 205 trials for each temperature condition, and an experiment with 10000 trials for each temperature condition.

JMP Pro 12.0 was used to conduct our statistical analysis (SAS Institute Inc., 2015). Due to the fact that we utilized CRN, a paired-*t* test was used for hypothesis testing unless otherwise noted (Law, 2015). Results for paired-*t* tests are reported as confidence intervals as suggested by Law (2015). For all tests, we utilize an alpha level (α) of 0.05. We ensured that we met assumptions and conditions for having paired data, independence between the mean values found for each temperature condition, randomization, and normality of the "*differences*" (De Veaux et al., 2012, p. 653). We must also note that although we compare the outcome of the hot temperature condition RNG sample to the mild temperature condition normal distribution to derive the IED danger area offset size, we preserve independence because we compare the RNG sample output against the ideal specified mild temperature condition normal distribution and not against the mild temperature condition normal distribution output from the RNG.

B.   HYPOTHESIS RESULTS

We now present the findings for our testing of each of our hypotheses related to our model implementation. Hypothesis $1A_0$ and $1B_0$ deal primarily with the RNG-generated sample distributions for each temperature condition in both the 205 and 10000 trial experiments. Hypothesis 2_0 addresses testing of the mean difference in IED danger area offset radii between the agent paths for each temperature condition, while

Hypothesis 3_0 tests the mean difference in path lengths between each temperature condition.

The run time for the 205 trial experiment for both the hot and mild temperature conditions took thirty-seven seconds. The run time for the 10000 trial experiment in the hot condition took nineteen minutes and seven seconds, while the run time for the mild temperature condition took seventeen minutes and thirty-five seconds.

1.   **Hypothesis 1A$_0$: Testing of Mean Difference for Temperature Condition Normal Distribution Random Number Generated Samples**

Our first hypothesis test deals with the mean difference between the random number generator (RNG)-generated temperature conditions normal distribution samples to determine if the two temperature conditions exhibited statistically significant differences after deriving their means and standard deviations from Hancock et al.'s (2007) meta-analysis.

a)   Results for n = 205 Trials for Each Temperature Condition

To test the null hypothesis that the mean difference between the RNG-generated temperature conditions normal distribution samples equaled zero, we conducted a paired-*t* test where: $\bar{y}_M = 3.00$, mild temperature condition standard deviation (s_M) = 0.66 and $\bar{y}_H$ = 2.84, hot temperature condition standard deviation (s_H) = 0.88. Results indicate that we can be ninety-five percent confident that the true mean difference between the RNG-generated temperature condition normal distribution samples is between 0.13 and 0.19. Therefore, we reject the null hypothesis. Figure 27 displays the histogram of the differences between the temperature condition normal distribution RNG-generated samples (difference equals mild temperature condition sample minus hot temperature condition sample; counts of data within a given bin are displayed above the bar).
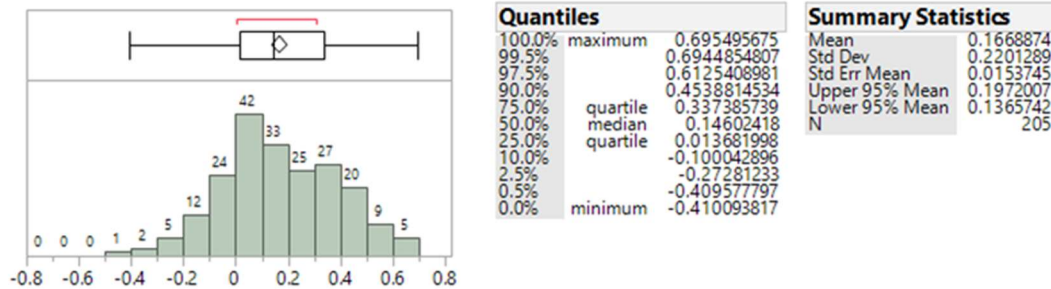
84

Quantiles			Summary Statistics	
100.0%	maximum	0.695495675	Mean	0.1668874
99.5%		0.6944854807	Std Dev	0.2201289
97.5%		0.6125408981	Std Err Mean	0.0153745
90.0%		0.4538814534	Upper 95% Mean	0.1972007
75.0%	quartile	0.337385739	Lower 95% Mean	0.1365742
50.0%	median	0.14602418	N	205
25.0%	quartile	0.013681998		
10.0%		-0.100042896		
2.5%		-0.27281233		
0.5%		-0.409577797		
0.0%	minimum	-0.410093817		

Figure 27.   Histogram, Boxplot, and Summary Statistics of the Differences
between Mild Temperature Condition and Hot Temperature Condition
RNG-Generated Samples (n = 205)

b)   *Results for n = 10000 Trials for Each Temperature Condition*

Our results for a paired-t test for the 10000 trials experiment with $\bar{y}_M$ = 2.98, mild temperature condition standard deviation (s_M) = 0.60 and $\bar{y}_H$ = 2.81, hot temperature condition standard deviation (s_H) = 0.80 tested the same hypothesis. These results indicate that we can be ninety-five percent confident that the true mean difference between the RNG-generated temperature condition's normal distributions is between 0.170 and 0.178. Therefore, we reject the null hypothesis that the mean difference between the mild and hot temperature condition distribution samples equaled zero. Figure 28 displays the histogram of the differences (difference equals mild temperature condition sample minus hot temperature condition sample).
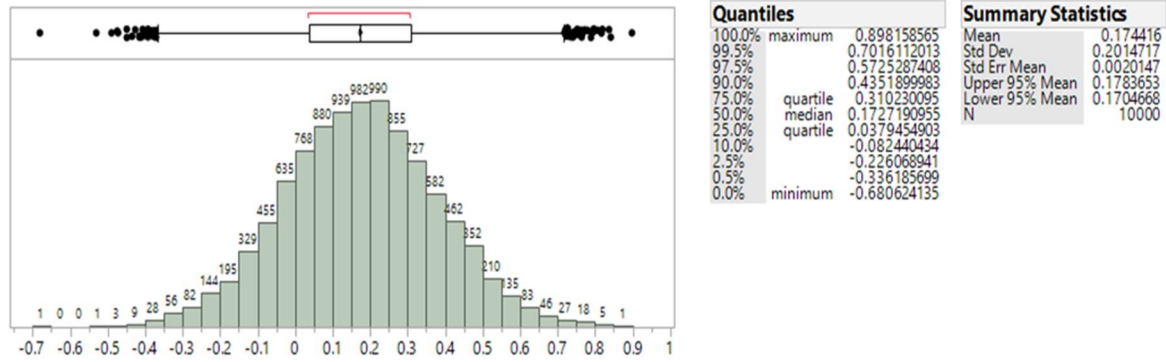
Figure 28.   Histogram, Boxplot, and Summary Statistics of the Differences between Mild Temperature Condition and Hot Temperature Condition RNG-Generated Samples (n = 10000)

2.   Hypothesis 1B$_0$: Testing of Temperature Condition Distribution Sampling

It was critical to determine the accuracy with which the RNG sampled from the normal distribution of each temperature condition, since the accuracy of this sampling would likely have an effect on the chosen value for the IED danger area radius (i.e., *randomNumberDASize*). Utilizing a one sample *t*-test, we conducted testing of our first null hypothesis: that there is no difference between the mean of the respective temperature condition's RNG-generated sample normal distribution and the specified mean for the temperature condition's normal distribution. Essentially, we tested the accuracy of the RNG normal distribution sampling method (Box-Muller algorithm) offered by the *MathNetForUnity* package. Additionally, we ensured the assumptions and conditions for a one-sample *t*-test of independence, normality, and randomization were met as outlined by DeVeaux, Velleman, & Bock (2012).

a)   *Results for n = 205 Trials for Each Temperature Condition*

The histograms and box plots for the RNG-generated sample output for the hot temperature condition (Figure 29) and mild temperature condition (Figure 30) appear generally normal.
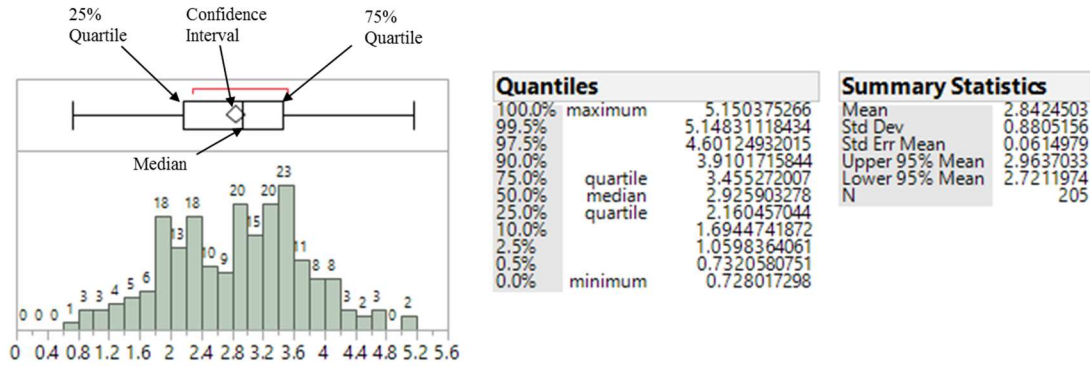
86

Figure 29.   Hot Temperature Condition RNG Sample Histogram, Box Plot &
Summary Statistics (n = 205)

Figure 30.   Mild Temperature Condition RNG Sample Histogram, Box Plot, and
Summary Statistics (n = 205)

A one-sample t-test for the hot temperature condition ($\bar{y}_H$ = 2.84, s_H = 0.88) indicates that there is no difference between the RNG-generated hot temperature condition normal distribution sample and the specified hot temperature condition normal distribution, $t(204) = 0.20$, $p = 0.83$. Therefore, we fail to reject the null hypothesis, meaning we have an accurate enough RNG-generated hot temperature condition normal distribution sample. Similarly, a one-sample t-test for the mild temperature condition ($\bar{y}_M$ = 3.00, s_M = 0.66) indicates that there is no difference between the RNG-generated mild temperature condition normal distribution sample and the specified mild temperature condition normal distribution, $t(204) = 0.20$, $p = 0.83$. Again, we fail to reject the null hypothesis.

We also conducted goodness of fit testing[20] of the RNG-generated distributions in accordance with Law (2015) to ensure they comprised a normal distribution. We utilized the Shapiro-Wilk Test (Shapiro & Wilk, 1965) to test both the hot and mild temperature condition RNG-generated distribution samples. The null hypothesis for this test states that the outcome distribution's samples come from the normal distribution (SAS Institute Inc., 2015). For the hot temperature condition, we fail to reject the null hypothesis ($W = 0.99$, $p = .26$). We also fail to reject the null hypothesis for the mild temperature condition's distribution ($W = 0.99$, $p = 0.26$). Therefore, we may proceed assuming that both RNG distributions are normal.

b)   *Results for n = 10000 Trials for Each Temperature Condition*

The histograms and box plots for each temperature condition's RNG-generated distribution outcomes appear in Figure 31 and Figure 32.

Figure 31.   Hot Temperature Condition RNG Sample Histogram, Box Plot and Summary Statistics (n = 10000)

---

[20] The goodness of fit tests we use here are empirical distribution function tests, not Chi-square tests (JMP 12 Online Documentation, n.d.).
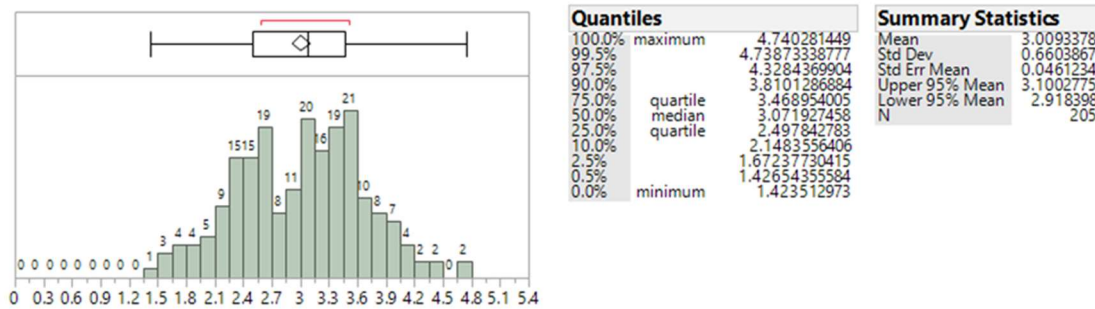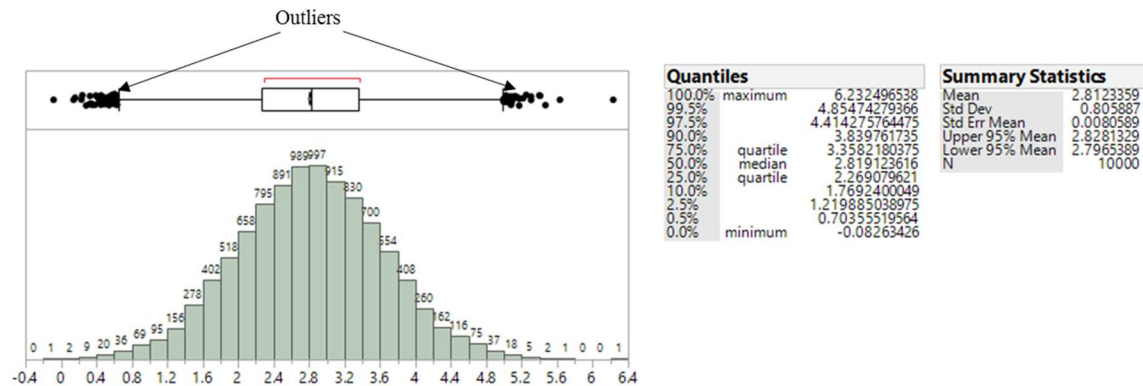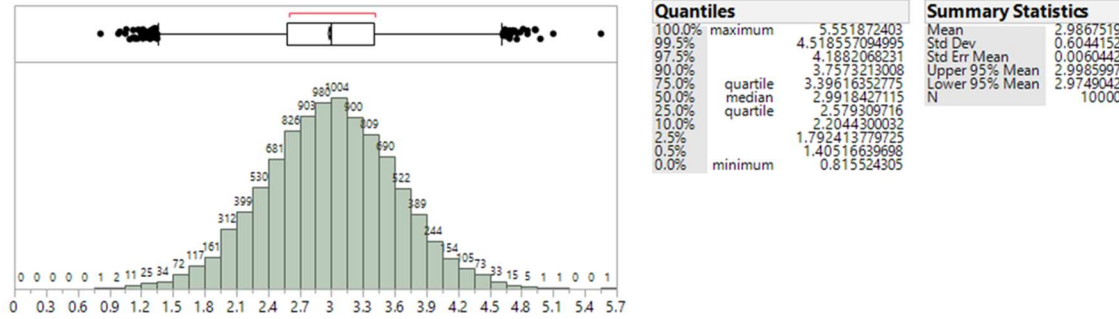
Figure 32.   Mild Temperature Condition RNG Sample Histogram, Box Plot, and Summary Statistics (n = 10000)

A one-sample *t*-test for the hot temperature condition ($\bar{y}_H$ = 2.81, s_H = 0.80) indicates that there is a difference between the RNG-generated hot temperature condition normal distribution sample and the specified hot temperature condition normal distribution, $t(9999) = -2.19$, $p = 0.02$. Therefore, we reject the null hypothesis. Similarly, a one-sample t-test for the mild temperature condition ($\bar{y}_M$ = 2.98, s_M = 0.60) indicates that there is a difference between the RNG-generated mild temperature condition normal distribution sample and the specified mild temperature condition normal distribution, $t(9999) = -2.19$, $p = 0.02$. Again, the null hypothesis is rejected.

We utilized the Kolmogorov-Smirnov-Lilliefors (KSL) test (Lilliefors, 1967) to determine if the hot and mild temperature condition RNG-generated distributions again fit a normal distribution (Law, 2015). This test utilizes the same null hypothesis as the Shapiro-Wilk test, but serves as *JMP 12.0's* goodness of fit testing mechanism for sample sizes over 2000 (JMP 12 Online Documentation, n.d.). For the hot temperature condition, we fail to reject the null hypothesis ($D = 0.004$, $p > 0.15$). We also fail to reject the null hypothesis for the mild temperature condition's distribution ($D = 0.004$, $p > 0.15$). Therefore, we may proceed assuming that both RNG distributions are normal.

3. Hypothesis 2_0: Testing of Mean of Differences for IED Danger Area Radius

We now provide the results from testing the mean difference between each temperature condition sample's IED Danger Area Radius sizes (essentially the value assigned to *randomNumberDASize* for each temperature condition).

a) Results for n = 205 Trials for Each Temperature Condition

To test the null hypothesis that the mean difference between each temperature condition sample's IED Danger Area Radius sizes (i.e., *randomNumberDASize*) equaled zero, we conducted a paired-t test where: $\bar{y}_M$ for danger area radius = 195.21, s_M for danger area radius = 66.46 and $\bar{y}_H$ for danger area radius = 180.14, s_H for danger area radius = 84.14. Results indicate that we can be ninety-five percent confidence that the true mean difference between each temperature condition sample's danger area radius sizes is between 12.02 and 18.12. Therefore, we reject the null hypothesis. Figure 33 displays the histogram of the differences between each temperature condition's IED Danger Area Radius sizes (difference equals mild temperature condition minus hot temperature condition).
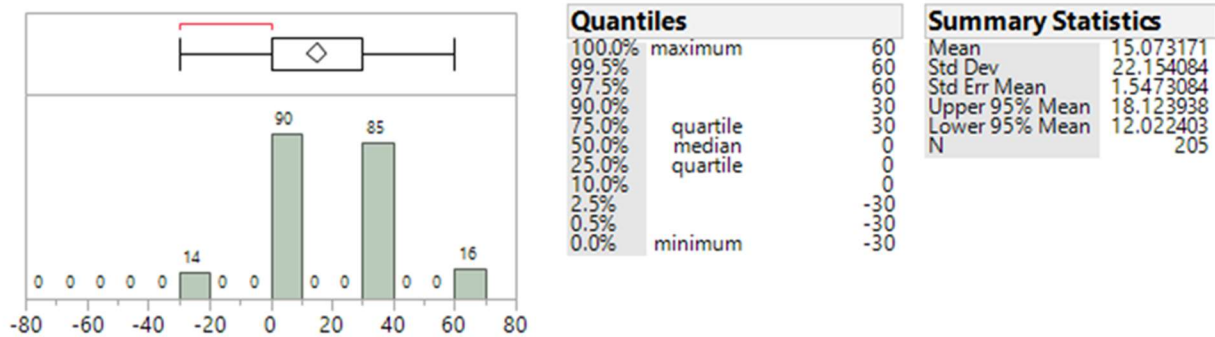

Figure 33.   Histogram, Boxplot, and Summary Statistics of the Differences between Mild Temperature Condition and Hot Temperature Condition IED Danger Area Radius Size (n = 205)

We tested the same hypothesis in an experiment with 10000 trials for each temperature condition. We conducted a paired-t test where: $\bar{y}_M$ for danger area radius = 193.64, s_M for danger area radius = 60.89 and $\bar{y}_H$ for danger area radius = 176.97, s_H for danger area radius = 77.82. We can be ninety-five percent confident that the true mean difference between each temperature condition sample's danger area radius sizes is between 16.23 and 17.09. Therefore, we reject the null hypothesis. Figure 34 displays the histogram of the differences between each temperature condition sample's IED Danger Area Radius sizes (difference equals mild temperature condition minus hot temperature condition).

Figure 34.   Histogram, Boxplot, and Summary Statistics of the Differences between Mild Temperature Condition and Hot Temperature Condition IED Danger Area Radius Size (n = 10000)

4.   Hypothesis 3_0: Testing of Mean of Differences for Path Length

Our final hypothesis tests the mean difference between each temperature condition sample's path lengths (in meters).

*a)   **Results for n = 205 Trials for each temperature condition***

To test the null hypothesis that the mean difference between each temperature condition sample's path lengths equaled zero, we conducted a paired-*t* test where: $\bar{y}_M$ for path length = 806.73, s_M for path length = 107.33 and $\bar{y}_H$ for path length = 798.74, s_H for path length = 132.00. These results suggest that we can be ninety-five percent confident

that the true mean difference between each temperature condition sample's path lengths is between 3.35 and 12.62 meters. Therefore, we reject the null hypothesis. Figure 35 displays the histogram of the differences between each temperature condition sample's path lengths (difference equals mild temperature condition minus hot temperature condition).

Quantiles		
100.0%	maximum	62.6673
99.5%		62.6673
97.5%		49.7058
90.0%		34.79402
75.0%	quartile	24.8529
50.0%	median	0
25.0%	quartile	0
10.0%		-2.6674
2.5%		-107.107
0.5%		-119.167
0.0%	minimum	-119.167

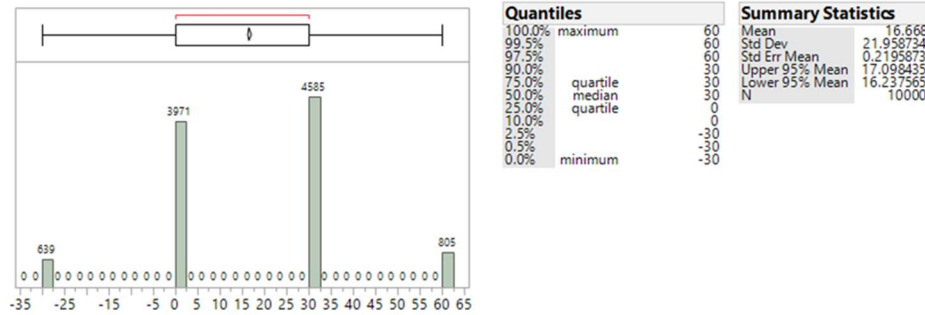| Summary Statistics	
Mean	7.9895541
Std Dev	33.637719
Std Err Mean	2.3493602
Upper 95% Mean	12.621696
Lower 95% Mean	3.3574125
N	205

Figure 35.   Histogram, Boxplot, and Summary Statistics of the Differences between Mild Temperature Condition and Hot Temperature Condition Path Length (in meters; n = 205)

b)   *Results for n = 10000 Trials for Each Temperature Condition*

Our results for the 10000 trial for each temperature condition experiment utilizing a paired-*t* test where: $\bar{y}_M$ for path length = 800.10, s_M for path length = 97.89 and $\bar{y}_H$ for path length = 789.92, s_H for path length = 121.30. We can be ninety-five percent confident that the true mean difference between each temperature condition sample's path lengths is between 9.54 and 10.82 meters. Therefore, we reject the null hypothesis. Figure 36 displays the histogram of the differences between each temperature condition sample's path lengths (difference equals mild temperature condition minus hot temperature condition).
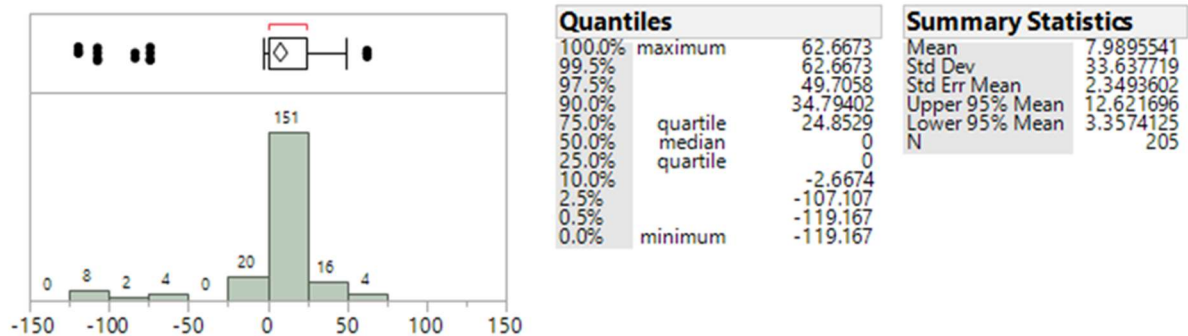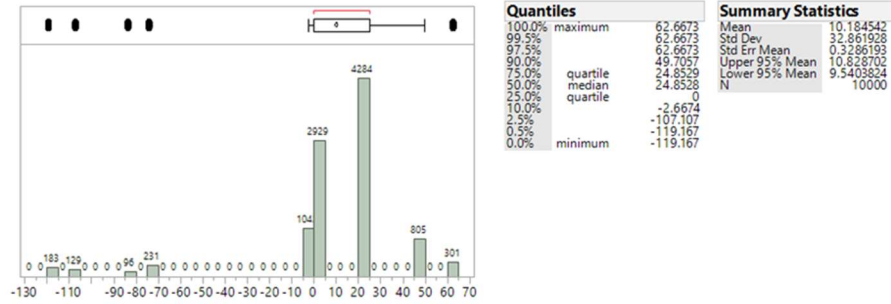
Figure 36.   Histogram, Boxplot, and Summary Statistics of the Differences
between Mild Temperature Condition and Hot Temperature Condition
Path Length (in meters; n = 10000)

THIS PAGE INTENTIONALLY LEFT BLANK.

V. DISCUSSION

A. ANALYSIS OF RESULTS

The results presented confirm that we can create agent performance exhibiting more decremented path finding around an IED danger area in hot temperature conditions relative to mild temperature conditions. Our findings show that in mild temperature conditions the agent travels paths demonstrating greater offset around an IED danger area than path offsets in hot temperature conditions. Furthermore, the agent paths travel a longer distance in mild temperature conditions than in hot temperature conditions. Another key contribution from our experiment demonstrates that we can demonstrate greater variability of performance in the hot temperature condition compared to the mild temperature condition in accordance with Hancock et al.'s (2007) observations.

In general, our mechanism for demonstrating greater performance decrement in hot temperature environments displays the desired behavior of smaller offsets from the IED danger area. This smaller offset provides a demonstration of decremented attention and vigilance to the IED danger area in a hot temperature environment. Following naturally from this small offset from the IED danger area, the path lengths in the hot temperature environment have a shorter length than the paths in the mild temperature environment. Hypothesis 2_0 for testing the mean difference in IED danger area radius size between the two temperature conditions proves most important, since the fundamental decision of our model's algorithm hinges on the assignment of the value for this radius size. Despite the different methods for calculating the "g" cost for both the hot and mild temperature conditions, the path length appears to be largely determined by the IED danger area radius size. Simply put, paths may look slightly different from the hot condition to the mild condition, but they will have the same path lengths if they have the same IED danger area radius size.

The larger value exhibited in the standard deviations for IED danger area size and path length for the hot temperature condition versus those for the mild temperature condition serves as an indicator that our model simulates Hancock et al.'s (2007)

discussion of the importance of variability measures in assessing the effects of heat stress. Furthermore, simulating this property allows us to put forth our model as one that promotes the importance assigned by the authors of 85ºF ET/87.4ºF WBGT as a "useful theoretical and practical threshold for the parsing of heat stressor effects" (p. 860).

We checked our output results for the positive correlation and reduced variance for the mean difference values as described by Law (2015). Our findings indicate that positive correlation did occur when conducting paired-t tests for Hypothesis $1A_0$, 2_0, and 3_0 (of note, the paired-t test for Hypothesis $1A_0$ indicated a correlation of 1.00 for both the 205 and 10000 trial experiments between the RNG-generated distribution samples for the mild and hot temperature conditions). Ultimately, we find the best evidence of the tighter confidence intervals and reduced variance of the mean differences from the CRN technique as described by Law (2015) when we look at the much smaller confidence intervals produced by the 10000 trial experiment compared to the confidence intervals produced by the 205 trial experiment.

We also note that the confidence intervals found in our test results for both the 205 and 10000 trial experiments showcase the desired behavior of performance decrement in a hot temperature environment without producing absurdly large values for the mean differences between the two temperature conditions. As discussed previously, we desired a model that would allow for behavior that proves realistic and believable in that ground combatants can make both good and bad decisions in mild and hot temperature conditions. Our model allows for this kind of behavior to occur, while still exhibiting our fundamental goal of exhibiting more decremented vigilance and attention performance in hot temperature conditions versus mild temperature conditions. This also speaks to the fact that outlier samples from the RNG-generated temperature condition distributions prove desirable and expected in that they promote this kind of realistic positive and negative behavior regardless of temperature condition.

Analyzing the results of Hypothesis $1A_0$ demonstrate statistically significant differences between the hot and mild temperature condition RNG-generated distribution samples for both the 205 trial and 10000 trial experiments. This aids in confirming that our technique described in Chapter III for assignment of values for the mean and standard

deviation of the hot and mild temperature conditions in accordance with Hancock et al. (2007) proved adequate for simulating the results of their meta-analysis.

Our results for testing Hypothesis $1B_0$ for the 205 trial experiment displays RNG-generated distribution samples in both temperature conditions that have a mean that proves statistically similar to the originally intended mean specified to the RNG. However, Hypothesis $1B_0$ demonstrates some interesting findings with respect to the 10000 trial experiment. Here we find for both temperature conditions that the RNG-generated distribution sample proves statistically different from the desired mean specified to the RNG. We can also see that for both the hot and mild temperature conditions the mean of the RNG-generated sample is less than the desired means by a small amount. Interestingly, the standard deviations for both temperature conditions appear quite exact to the standard deviations specified to the RNG. It remains unclear why the RNG produced distribution samples with statistically different means from the specified mean for both temperature conditions, and why both means differ from the specified mean by a similar amount.

The Box-Muller algorithm used for generating our RNG normal distribution samples may be the culprit of the inaccuracy problem. Multiple sources describe some problems with the Box-Muller algorithm, and recommend against its use—especially with certain types of random number generators known as linear congruential generators (Law, 2015; Neave, 1973; Potuznik & Hinow, 1997). Law (2015) specifically warns that the random numbers coming from the Box-Muller algorithm used with this type of generator could result in a spiral pattern when plotted, "rather than being truly independently normally distributed" (p. 457). Potuznik & Hinow (1997) give examples of what these plots could look like (see Figure 37 for an example of a Box-Muller output from a linear congruential generator). The plots consist of points representing consecutive pairs of normal distribution sampled numbers, i.e., the first point represents (X_1, X_2), the next point represents (X_3, X_4), etc., where each X is a number sampled from the normal distribution by the Box-Muller algorithm (Law, 2015; Potuznik & Hinow, 1997).
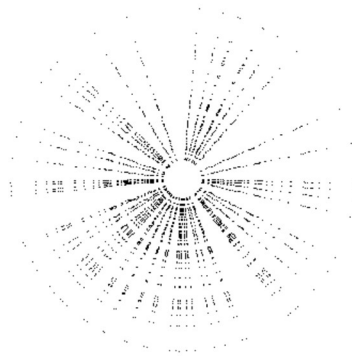
Figure 37.   Box-Muller Algorithm Output for Linear Congruential Generator.
Source: Potuznik and Hinow (1997).

Afflerbach & Wenzel (1988) describe certain circumstances where using the Box-Muller algorithm may not prove detrimental to use, and Law (2015) cautiously admits that other types of random number generators could still allow for using the Box-Muller algorithm. Figure 38 displays an example where an issue in the Box-Muller output did not appear to arise as a result of using an algorithm by Wichmann & Hill (1987), described by Potuznik & Hinow (1997) as a "linear combination of three linear congruency methods [linear congruential generators]" (p. 3).
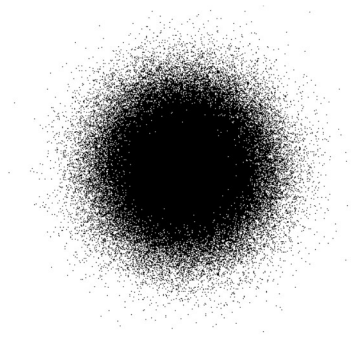
Figure 38.   Box-Muller Algorithm Output for Wichmann & Hill Algorithm.
Source: Potuznik and Hinow (1997).

Law (2015) does not mention whether the Mersenne Twister random number generator, a type of feedback shift register generator, could suffice to prevent the spiral

pattern issue. However, due to the results of our goodness of fit tests indicating that our RNG-generated distribution samples for both conditions came from the Normal distribution, and plots of our RNG-generated number pairs for the hot and mild temperature conditions (see Figure 39 through Figure 42) showing no spiral patterns, these reservations may prove of little importance to our model.

Figure 39.   Mild Temperature Condition Plot of Normal Distribution Sample
Pairs (n = 205; 102 pairs).
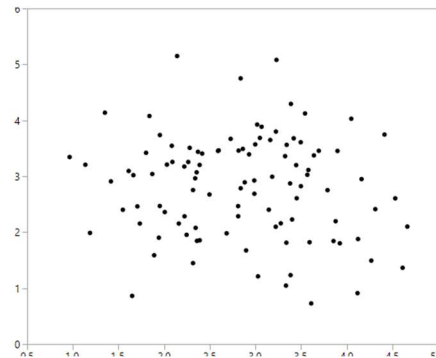
Figure 40.   Hot Temperature Condition Plot of Normal Distribution Sample
Pairs (n = 205; 102 pairs).
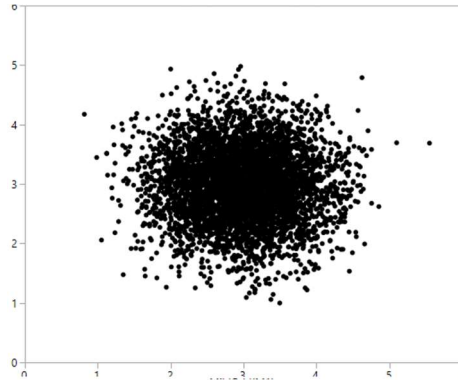
Figure 41.   Mild Temperature Condition Plot of Normal Distribution Sample
Pairs (n = 10000; 5000 pairs).
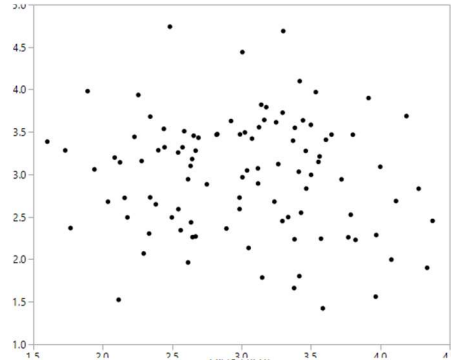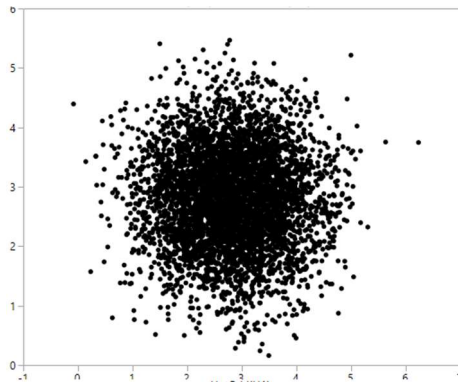

Figure 42.   Hot Temperature Condition Plot of Normal Distribution Sample
Pairs (n = 10000; 5000 pairs).

Ultimately, Law (2015) does argue for utilizing an "algorithm that results in random variates with exactly the desired distribution" (p. 427). However, Law (2015) also puts forth that in picking a type of probability distribution that "provides the best fit" of the data, we are also "trying to…determine a distribution that is accurate enough for the intended purposes of the model" (p. 335). Law (2015) focuses on simulation models outside of the realm of training simulations. Since our model is designed to improve the fidelity of decremented behavior in hot temperatures for trainees—which we have demonstrated a mechanism for—we propose that the accuracy issues of the sample distributions identified in the 10000 trial experiment, and concerns over use of the Box-Muller algorithm appear to be inconsequential.

B. LIMITATIONS

A primary limitation of the model stems from the source of the model's statistical data from the meta-analysis done by Hancock et al. (2007). Ideally, the data used for our model implementation would come from studies of military personnel executing the types of military tasks modeled in our proof-of-concept model implementation, but the meta-analysis draws from studies across numerous domains. Given limited resources for a human subjects experiment, the meta-analysis offered the best breakdown of the data according to perceptual types of tasks in a hot, humid environment. However, Hancock et al. (2007) does call attention to the general limitation of the meta-analysis in that it does not account for all of the "moderator variables" likely to be present in the studies utilized (p. 856). Lowe et al. (2007) provides a number of possible data sources in the literature from studies of military personnel, but we believe that isolating data measurements focused on vigilance and attention types of tasks would prove too difficult given the inability to control and account for a number of factors present in these kinds of studies. Thus, the Hancock et al. (2007) meta-analysis proved sufficient for our model.

Similarly, we make numerous assumptions within our model to quantify vigilance and attention-type task performance in hot or mild temperature environments into a path that hinges largely upon the value selected for *randomNumberDASize*. Our ability to do this may indirectly fulfill the observations by Logan and Alechina (1998) and Ritter et al. (2003) about the potential impossibility of boiling a real problem, or human behavior, down to one parameter. A human subjects experiment would allow us to determine if such assumptions in our model actually proved detrimental to realism.

Hancock et al.'s (2007) meta-analysis required that study temperatures could be utilized to calculate WBGTs, and we presented a mechanism for visually depicting the WBGT in the environment to other agents via indicators of the environmental WBGT flag conditions. However, the Marine Corps order describing the flag condition guidelines states that wearing body armor adds 10°F to the WBGT value (Commandant of the Marine Corps, 2003). Our interpretation of Hancock et al.'s (2007) meta-analysis does not include this additional factor, when such a temperature addition would prove highly relevant given the use of body armor by today's Soldiers and Marines.

Hancock et al. (2007) conducted their task breakdowns according to the ET scale. Even though the authors required studies to contain enough information for WBGT calculation to become a part of the meta-analysis, Hancock and Vasmatzidis (2003) identify some issues of incompleteness of information in the ET conversion to WBGT. Essentially, ET constitutes a "subjective scale" lacking what the WBGT scale includes in a mechanism for "directly specifying the intensity of the radiant heat stress element" (Hancock & Vasmatzidis, 2003, p. 369). Further, the same WGBT temperature can be achieved with "different dry-bulb/relative humidity combinations" (p. 369). Our model does not account for these finer details and issues with respect to temperature.

Another limitation of the model focuses on the paths found by the agent. In our current implementation, the model's path allows travels around only one side of the IED danger area location. A possible solution for this issue would entail an additional decision point within the model to increase the cost of nodes on one side of the IED danger area every few path calculations. If done properly, this could encourage agent paths that travel to the opposite side of the IED danger area at random intervals.

Similarly, the agent's paths demonstrated more jaggedness at times than one would realistically expect a human path to travel. We did not implement any string pulling techniques to eliminate jaggedness, partly because *Apex Path's* built-in mechanism for doing so eliminated the desired offset path behavior around the IED danger area. Additional coding implementation could eliminate the jaggedness in paths, but we remained chiefly concerned with demonstrating overall decremented path finding behavior due to environmental heat.

C.   FUTURE WORK

We recommend several areas for future work that would greatly improve our model proof-of-concept. A human subjects experiment to provide a basis for validating our model comprises the most important area of future work largely due to the number of assumptions and missing variables in our model. Another key area of future work would entail creating and implementing our model proof-of-concept in a constructive simulation

such as COMBATXXI, aligning with recommendations by Blais (2016) to improve human behavior models within constructive simulation.

Additional areas of future work that would make the model more useful would entail the implementation of greater variability in IED danger area locations, as well as the incorporation of multiple danger areas for the agent to navigate around. Our model maintains the same location for the IED danger area, as well as the same start and destination points for the agent. Future work could vary these locations at the outset of each trial to test the flexibility of the environmentally impacted A* algorithm in these varied situations.

Computer vision techniques and algorithms for simulation target acquisition could provide a domain for future work by using our translation of Hancock et al.'s (2007) findings into a means to influence the effectiveness of techniques described in Darken and Jones (2007). We could also see applying techniques used in this thesis to influence the effectiveness of visual search and threat location models described in Darken & Anderegg (2008) and Darken, McCue, and Guerrero (2010). Alt & Darken (2008) offer a fairly robust synthesis of models for modeling a "soldier's visual attention (search)…and target acquisition" (p. 1). The authors describe their model as including a mechanism for Endsley's "level 1" environmental perception aspect of situational awareness pertaining to their implementation of agent awareness of enemy threat location and an understanding of the simulation's terrain map (as cited in Alt & Darken, 2008, p. 8). Future work should also explore incorporating various aspects of this thesis into the portions of Alt & Darken's (2008) model, as well as the portion of their model dealing with selecting an action for the entity.

Incorporating multiple agents navigating as a team under environmental impacts on vigilance and attention is another logical extension of the work accomplished by this thesis in order to provide a solution for more realistic training to Marines and soldiers operating as a part of larger units. Such future work could build upon the work of Darken, McCue, & Guerrero (2010) who implement team and squad path finding with an A* algorithm.

D.   CONCLUSION

Ultimately, we have provided a model for virtual simulation entities based upon the environmental impacts of heat and humidity on human information processing. This model makes use of the A* path finding algorithm to demonstrate how heat and humidity can negatively impact the performance of an agent's vigilance and attention paid to their surrounding environment, resulting in decremented path finding performance of the agent as they attempt to navigate around an IED danger area location. We believe that including implementations similar to the one proposed in this thesis within military training simulations can greatly improve the accuracy of simulation agent behavior. More importantly, we believe that such enhanced accuracy could improve the training of military personnel utilizing the simulation so that they gain a better understanding of the potential issues that they may encounter on the battlefield while operating as a member of a unit.

SUPPLEMENTAL

This supplemental describes the functionality and model execution within the Improved Performance Research Integration Tool (IMPRINT). This document has a restricted distribution. Those interested in accessing this material should contact the NPS Dudley Knox Library.

THIS PAGE INTENTIONALLY LEFT BLANK.

LIST OF REFERENCES

Abdellaoui, N., Taylor, A., & Parkinson, G. (2009). *Comparative analysis of computer generated forces' artificial intelligence* (Report No. RTO-MP-MSG-069). Retrieved from http://www.dtic.mil/dtic/tr/fulltext/u2/a567877.pdf

Afflerbach, L., & Wenzel, K. (1988). Normal random numbers lying on spirals and clubs. *Statistical Papers*, *29*(1), 237–244. https://doi.org/10.1007/BF02924529

Allender, L. (2000). Modeling human performance: Impacting system design, performance, and cost. *Simulation Series*, *32*(3), 139–144.

Allender, L., Kelley, T., Lockett, J., & Archer, S. (2005). Human performance modeling in the Army: A long and winding road. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *49*(12), 1191–1195, doi: 10.1177/154193120504901220

Alt, J. K., & Darken, C. J. (2008). A reference model of soldier attention and behavior. In *Proceedings of Behavior Representation in Modeling and Simulation (BRIMS) 2008*. Providence, RI. Retrieved from http://www.dtic.mil/dtic/tr/fulltext/u2/a534859.pdf

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, NJ: Lawrence Erlbaum Associates.

Apex Game Tools. (2016). Apex Path (Version 2.4.5) [Computer software]. Retrieved from https://www.assetstore.unity3d.com/en/#!/

Apex Game Tools. (n.d.). Apex Path Video Tutorials [YouTube Videos]. Retrieved from http://apexgametools.com/learn/apex-path-documentation/apex-path-tutorials/

Aquarius Max. (2016). Modular Prison Fortress (Version 1.3) [Unity Package]. Retrieved from https://www.assetstore.unity3d.com/en/#!/

Army gaming releases newest training product. (2014, April 10). *FtLeavenworthLamp.com*. Retrieved from http://www.ftleavenworthlamp.com/article/20140410/News/140419879/?tag=1

Bailey, M. P. (2002). The deployable virtual training environment. *Calhoun: The NPS Institutional Archive*. Retrieved from http://calhoun.nps.edu/bitstream/handle/10945/40350/IITSEC.bailey.armstrong-final.pdf?sequence=1

Best, B. J., Lebiere, C., & Scarpinatto, K. C. (2002). Modeling synthetic opponents in MOUT training simulations using the ACT-R cognitive architecture. In *Proceedings of the 11th Conference on Computer Generated Forces and Behavior Representation*, *7,* 33–44. Retrieved from http://adcogsys.com/pubs/11TH-CGF-056.pdf

Bismur Studios Ltd. (2017a). Math.NET for Unity (Version 1.0.1) [Unity Package]. Retrieved from https://www.assetstore.unity3d.com/en/#!/

Bismur Studios Ltd. (2017b). Math.NET for Unity Documentation (Version 1.0.1) [Doxygen 1.8.11 Documentation]. Retrieved from Math.NET for Unity Package

Blais, C. (2016). Challenges in representing human-robot teams in combat simulations. In J. Hodicky (Ed.), *Proceedings of the 2016 Modeling and Simulation for Autonomous Systems (MESAS) Workshop*. NATO M&S Center of Excellence.

Bohemia Interactive Simulations. (2015). VBS3 manuals [Online Documentation]. Retrieved from https://manuals.bisimulations.com/vbs3/3-7/manuals/Default.htm

Borgman, A. (2007, March). Human behavior in the Infantry Warrior Simulation (IWARS) [PowerPoint]. Presented at *16th BRIMS Conference*. Retrieved from https://orsagouge.pbworks.com/f/07-BRIMS-060.ppt

Borkman, S., Verdesca, M., & Watkins, J. (2012). *Algorithms for ground soldier based simulations and decision support applications phase I* (Report No. NATICK/TR-12/019) [Technical Report]. Retrieved from http://www.dtic.mil/dtic/tr/fulltext/u2/a560713.pdf

Box, G. E. P., & Muller, M. E. (1958). A note on the generation of random normal deviates. *The Annals of Mathematical Statistics*, 29(2), 610–611. Retrieved from http://www.jstor.org/stable/2237361

Carroll, M., Padron, C. K., Quinn, S., Surpris, G., Winslow, B., & Viklund, E. (2015). Effectiveness of process level feedback at training tactical decision making. Paper No: 15201. 1–12. Presented at the *Interservice/Industry Training Simulations, and Education Conference (I/ITSEC) 2015*. Retrieved from http://www.iitsecdocs.com

Chun, M. M., Golomb, J. D., & Turk-Browne, N. B. (2011). A taxonomy of external and internal attention. *Annual Review of Psychology*, *62*, 73–101.

Commandant of the Marine Corps. (2003, July 23). Marine Corps Order 6200.1E W/ CH 1. *Headquarter, United States Marine Corps*. Retrieved from http://www.marines.mil/Portals/59/Publications/MCO%206200.1E%20W%20CH%201.pdf

Curthoys, K. (2014, December 8). Personalized approach to Army avatars. *DefenseNews.com*. Retrieved from http://www.defensenews.com/story/defense/international/americas/2014/12/08/personalized-approach-to-army-avatars-/20105585/

Dancy, C.L. (2013). ACT-RΦ: A cognitive architecture with physiology and affect. In *Biologically Inspired Cognitive Architectures* (Vol. 1, pp. 40–45). doi: 10.1016/j.bica.2013.07.008

Dancy, C. L., Ritter, F. E., & Berry, K. (2012). Towards adding a physiological substrate to ACT-R. *Proceedings of the 21st Conference on Behavior Representation in Modeling and Simulation*, 78–85. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.226.1571&rep=rep1&type=pdf

Darken, C. (2004). Visibility and concealment algorithms for 3D simulations. In *Proceedings of Behavior Representation in Modeling and Simulation*. Arlington, VA. Retrieved from http://www.dtic.mil/dtic/tr/fulltext/u2/a574810.pdf

Darken, C. (2016). *MV4025: Cognitive and Behavioral Modeling for Simulations*. Naval Postgraduate School.

Darken, C.J., & Anderegg, B. G. (2008). Particle filters and simulacra for more realistic opponent tracking. In S. Rabin (Ed.), *Game AI Programming Wisdom 4*. Boston: Course Technology.

Darken, C.J., & Jones, B. E. (2007). Computer graphics-based target detection for synthetic soldiers. In *Proceedings of Behavior Representation in Modeling and Simulation (BRIMS 2007)*. Norfolk, VA. Retrieved from http://www.dtic.mil/dtic/tr/fulltext/u2/a551431.pdf

Darken, C.J., McCue, D., & Guerrero, M. (2010). Realistic fireteam movement in urban environments. *Proceedings of the Sixth Artificial Intelligence for Interactive Digital Entertainment Conference (AIIDE 2010)*. Retrieved from http://www.dtic.mil

De Veaux, R. D., Velleman, P. F., & Bock, D. E. (2012). *Intro Stats* (3rd ed.). Boston: Addison-Wesley.

Edwards, D. J., Grodevant, N. W., Lee, P. J., & Peralta, J. B. (2007). DNA-MAN: Dynamic Natural Attributes for Synthetic Military Forces. In *2007 IEEE Systems and Information Engineering Design Symposium* (pp. 1–5). doi: 10.1109/SIEDS.2007.4374030

Eldridge, C., Andra, M., Kennedy, A., Peloquin, A., Rice, D., & Yuhaski, Jr., S. (2012). Task-related encumbrance methodology in Force-on-Force Agent-Based Simulation. In *Proceedings of the 2012 Symposium on Military Modeling and Simulation*. Society for Computer Simulation International. Retrieved from http://dl.acm.org/citation.cfm?id=2339452

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *37*(1), 32–64. Retrieved from https://doi.org/10.1518/001872095779049543

Evertsz, R., Pedrotti, M., Busetta, P., Acar, H., & Ritter, F. (2009). Populating VBS2 with realistic virtual actors. Presented at the *Proceedings of the 18th Conference on Behavior Representation in Modeling and Simulation*, Sundance, UT. Paper No: 09-BRIMS-004. 1–8. Retrieved from http://ritter.ist.psu.edu/papers/evertszpbar09.pdf

Evertsz, R., Ritter, F. E., Busetta, P., & Pedrotti, M. (2008). Realistic behaviour variation in a BDI-based cognitive architecture. *Proceedings of SimTecT 2008*. Retrieved from http://www.siaa.asn.au/library_simtect_2007.html

Fefferman, C. K., Diego, M., Gaughan, C., Samms, C., Borum, H., Clegg, J., McDonnell, J.S., & Leach, R. (2015). *A study in the implementation of a distributed soldier representation* (Report No. ARL-TR-6985). Aberdeen Proving Ground, MD: Army Research Lab. Retrieved from http://www.dtic.mil

Gillis, P. D., & Hursh, S. R. (1999). Using behavior moderators to influence CGF command entity effectiveness and performance. Presented at the *8th Conference on Computer Generated Forces and Behavioral Representation*. p. 237–251. Retrieved from http://www.ucdavis.edu

G-3 Operations and Training. (n.d.). BSC CACCTUS [Webpage]. Retrieved from http://www.29palms.marines.mil/Staff/G3-Operations-and-Training/MAGTFTC-Simulations/Combined-Arms-Training/CACCTUS/

Hancock, P. A., Ross, J. M., & Szalma, J. L. (2007). A meta-analysis of performance response under thermal stressors. *Human factors: The Journal of the Human Factors and Ergonomics Society*, *49*(5), 851–877. doi: 10.1518/001872007X230226

Hancock, P. A., & Vasmatzidis, I. (2003). Effects of heat stress on cognitive performance: the current state of knowledge. *International Journal of Hyperthermia*, *19*(3), 355–372. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.464.7830&rep=rep1&type=pdf

Hancock, P. A., & Warm, J. S. (1989). A dynamic model of stress and sustained attention. *Human Factors*, *31*(5), 519–537. doi: 10.1177/001872088903100503

Harder, B. R., & Darken, C. (2016). Automated fire support planning for combat simulations. *Proceedings of the International Conference on Social Computing, Behavioral-Cultural Modeling, & Prediction and Behavior Representation in Modeling and Simulation*, *7*. Retrieved from http://sbp-brims. org/2016/proceedings/IN

JMP 12 Online Documentation. (n.d.). Fit Distributions [Webpage]. Retrieved from http://www.jmp.com/support/help/Fit_Distributions.shtml#273232

Johnson, R. F., & Kobrick, J. L. (2001). Chapter 4: Psychological aspects of military performance in hot environments. In *Medical Aspects of Harsh Environments* (Vol. 1, pp. 135–159). Retrieved from: http://www.dtic.mil/dtic/tr/fulltext/u2/a409995.pdf

Juarez-Espinosa, O., & Gonzales, C. (2004). Situation awareness of commanders: A cognitive model. *2004 Conference on Proceedings of Behavior Representation in Modeling and Simulation*. Retrieved from http://repository.cmu.edu/cgi/viewcontent.cgi?article=1086&context=sds

Kelley, T. D., Patton, D. J., & Allender, L. (2001). Predicting situation awareness errors using cognitive modeling. In M. J. Smith, G. Salvendy, D. Harris, & R. J. Koubek (Eds.), *Proceedings of Human-Computer Interaction International 2001 Conference* (Vol. 1: Usability Evaluation and Interface Design: Cognitive Engineering, Intelligent Agents and Virtual Reality, pp. 1455–1459). Mahwah, NJ: Lawrence Erlbaum Associates.

Laird, J. E., Assanie, M., Bachelor, B., Benninghoff, N., Enam, S., Jones, B., … Wallace, S. (2002). *A test bed for developing intelligent synthetic characters* (AAAI Technical Report No. SS-02-01). Retrieved from http://www.aaai.org/Papers/Symposia/Spring/2002/SS-02-01/SS02-01-011.pdf

Laird, J. E., Newell, A., & Rosenbloom, P. S. (1987). Soar: An architecture for general intelligence. *Artificial intelligence*, *33*(1), 1–64. doi: https://doi.org/10.1016/0004-3702(87)90050-6

Law, A. M. (2015). *Simulation modeling and analysis* (5th ed.). New York, NY: McGraw-Hill Education.

Lidén, L. (2002). Strategic and tactical reasoning with waypoints. In *AI Game Programming Wisdom*. Charles River Media. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.132.5049

Lilliefors, H. W. (1967). On the Kolmogorov-Smirnov Test for normality with mean and variance unknown. *Journal of the American Statistical Association*, *62*(318), 399–402. https://doi.org/http://dx.doi.org/10.1080/01621459.1967.10482916

Logan, B., & Alechina, N. (1998). A* with bounded costs. *AAAI-98 Proceedings*, 444–449. Retrieved from http://www.aaai.org/Papers/AAAI/1998/AAAI98-063.pdf

Lopez, C. T. (2014, April 3). Latest "Virtual Battle Space" release adds realism to scenarios, avatars. *Army.mil*. Retrieved from https://www.army.mil/article/123316/Latest__Virtual_Battle_Space__release_adds_realism_to_scenarios__avatars

Lowe, M., Harris, W., Kane, R. L., Banderet, L., Levinson, D., & Reeves, D. (2007). Neuropsychological assessment in extreme environments. *Archives of Clinical Neuropsychology*, *22*, *Supplement 1*, 89–99. doi: https://doi.org/10.1016/j.acn.2006.10.010

Mahoney, C. R., Hirsch, E., Hasselquist, L., Lesher, L. L., & Lieberman, H. R. (2007). The effects of movement and physical exertion on soldier vigilance. *Aviation Space Environmental Medicine*, *78*(Supplement 1), B51–B57. Retrieved from http://www.ingentaconnect.com/content/asma/asem/2007/00000078/A00105s1/art00008

Marines.mil (n.d.). Flag Conditions [Webpage]. Retrieved from http://www.ready.marines.mil/Stay-Informed/Natural-Hazards/Extreme-Heat/Flag-Conditions/

Matsumoto, M., & Nishimura, T. (1998). Mersenne Twister: A 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation (TOMACS)-Special Issue on Uniform Random Number Generation*, *8*(1), 3–30. doi: https://doi.org/10.1145/272991.272995

Matthews, J., & van der Sterren, W. (2002). A* Explorer (Version 1.1 + tactical path-finding extensions) [Computer Program].

Microsoft Corporation. (2016). Visual Studio 2015 (Version 14.0.25425.01 Update 3) [Computer Program].

MTWS Program: Cole Engineering Services, Inc. MTWS post deployment software support & new requirements: MTWS user documentation, MTWS v3.5.5. (2016, November 17).

National Ground Intelligence Center. (n.d.). Improvised Explosive Device (IED) safe standoff distance cheat sheet. United States Army. Retrieved from http://www.respondersafety.com/Download.aspx?DownloadId=fda90981-f558-41f8-8bd2-395a03385258

Neave, H. R. (1973). On using the Box-Muller Transformation with multiplicative congruential pseudo-random number generators. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *22*(1), 92–97. https://doi.org/10.2307/2346308

North Atlantic Treaty Organization (NATO), T. G. 128. (2009). *Human behavior representation in constructive simulation* (Report No. TR-HFM-128) [Technical Report]. Retrieved from https://www.researchgate.net/profile/Martin_Castor2/publication/280945308_NATO_RTO_HFM_128_Human_Behavior_Representation_in_Constructive_Simulation/links/55cdf2f208ae6a8813815642.pdf

Oferei. (2014, January 12). Normal distribution random [Forum Post]. *answers.unity3d.com*. Retrieved from http://answers.unity3d.com/questions/421968/normal-distribution-random.html

Pandolf, K.B., Givoni, B., & Goldman, R.F. (1977). Predicting energy expenditure with loads while standing or walking very slowly. Journal of Applied Physiology: Respiratory, Environmental, and Exercise Physiology, *43*(4), 577–581. Retrieved from http://www.dtic.mil

Pandolf, K. B., Stroschein, L. A., Drolet, L. L., Gonzalez, R. R., & Sawka, M. N. (1985). *Prediction modeling of physiological responses and human performance in the heat* (pp. 1–33). Natick, MA. Retrieved from http://www.dtic.mil

Paulsen, M., Alicia, T. J., & Shrader, D. M. (2012). Modeling algorithms for predicting the effects of human performance in the presence of environmental stressors. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* Vol. *56*(1), 975–979. doi: 10.1177/1071181312561204

PEOStri. (n.d.). OneSAF [Webpage]. Retrieved from http://peostri.army.mil/one-semi-automated-forces-onesaf-

Pew, R. W., & Mavor, A. S. (1998). *Modeling human and organizational behavior: Application to military simulations*. National Academies Press. Retrieved from http://www.nap.edu/catalog/6173.html

Pfeiffer, K. D., & Tamash, T. (2014). Measuring the impact of natural environment representation on combat simulation outcomes. Paper No: 14305. 1–11. Presented at the *Interservice/Industry Training, Simulation, and Education Conference* (I/ITSEC) 2014. Retrieved from http://www.iitsecdocs.com

Pilcher, J. J., Nadler, E., & Busch, C. (2002). Effects of hot and cold temperature exposure on performance: a meta-analytic review. *Ergonomics*, *45*(10), 682–698. doi: http://dx.doi.org/10.1080/00140130210158419

Potuznik, M., & Hinow, P. (1997). Deterministic patterns in pseudorandom point sets. Presented at the *Workshop on Simulation, Dresden*. Retrieved from https://people.uwm.edu/hinow/files/2016/04/workshop_97-295aj7f.pdf

Program Manager Training Systems. (2016). PM TRASYS 2016 products & services catalog. Retrieved from http://www.marcorsyscom.marines.mil/Portals/105/PMTRASYS/Images/2016%20PROD%20AND%20SERVICES%20CAT%20-%20PM%20TRASYS.pdf?ver=2016-05-25-145144-750

Rabin, S. (2000a). A* aesthetic optimizations. In *Game Programming Gems* (pp. 264–271). Charles River Media, Inc.

Rabin, S. (2000b). A* speed optimizations. In *Game Programming Gems* (pp. 272–287). Charles River Media, Inc.

Reece, D. A., Kraus, M., & Dumanoir, P. (2000). Tactical movement planning for individual combatants. In *Proceedings of the 9th Conference on Computer Generated Forces and Behavioral Representation*. Retrieved from: https://skatgame.net/mburo/ps/cgf9-pathplanning.pdf

Reilly, W. S. N., Harper, K. A., & Marotta, S. (2007). Modeling concurrent, interacting behavior moderators for simulation-based acquisition tasks. *SpringSim*, *2*, 267–274. Retrieved from http://dl.acm.org/citation.cfm?id=1404854

Ritter, F. E., & Avraamides, M. N. (2000). Steps towards including behavior moderators in human performance models in synthetic environments (Report No: ACS 2000–1) [Technical Report]. Retrieved from http://s3.amazonaws.com/academia.edu.documents/32127139/rittera00.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1493850496&Signature=4k2NiGxa5Pihzg8JZg%2BhFsEKHN8%3D&response-content-disposition=inline%3B%20filename%3DSteps_Towards_Including_Behavior_Moderat.pdf

Ritter, F. E., & Norling, E. (2006). Including human variability in a cognitive architecture to improve team simulation. In R. Sun (Ed.), *Cognition and Multi-Agent Interaction: From Cognitive Modeling to Social Simulation* (pp. 417–427). Cambridge: Cambridge University Press.

Ritter, F. E., Shadbolt, N. R., Elliman, D., Young, R. M., Gobet, F., & Baxter, G. D. (2003). Techniques for modeling human performance in synthetic environments: A supplementary review, (Report No. HSIAC-SOAR-02-02). Retrieved from http://www.dtic.mil/dtic/tr/fulltext/u2/a488620.pdf

Russell, S. J., & Norvig, P. (2010). *Artificial intelligence: A modern approach* (Third). Upper Saddle River, New Jersey: Prentice Hall.

Santee, W. R., Reardon, M. J., & Pandolf, K. B. (2012). Chapter 2: Modeling the physiological and medical effects of exposure to environmental extremes. In *Military Quantitative Physiology: Problems and Concepts in Military Operational Medicine* (pp. 39–72). Fort Detrick, MD: Office of The Surgeon General United States Army.

SAS Institute Inc. (2015). JMP Pro (Version 12.0.1 64-bit) [Computer Program]. Retrieved from http://www.nps.edu/web/technology/software

Schleper, R., Gaughan, C., Kierzewski, M.O., Dunmire, C., Salvi, L., Fann, J., & Kellihan, B. (2010). Individual warfighter effectiveness and survivability in a CBRN threat environment. In *Proc. SPIE 7705, Modeling and Simulation for Defense Systems and Applications V,* 77050M. doi:10.1117/12.851903.

Schroth, T. F. (1989, March). *An introduction to human factors and combat models*. Naval Postgraduate School, Monterey, CA. Retrieved from http://hdl.handle.net/10945/27107

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4), 591–611. https://doi.org/10.2307/2333709

Silverman, B. G., Bharathy, G., O'Brien, K., & Cornwell, J. (2006). Human behavior models for agents in simulators and games: Part II: Gamebot engineering with PMFserv. *Presence*, *15*(2), 163–185. doi: 10.1162/pres.2006.15.2.163

Silverman, B. G., Johns, M., Cornwell, J., & O'Brien, K. (2006). Human behavior models for agents in simulators and games: Part I: Enabling science with PMFServ. *University of Pennsylvania Scholarly Commons*. Retrieved from http://repository.upenn.edu/ese_papers/296

skm. (2014, February 12). Image of HSV color wheel for openCV [Forum Post]. *stackoverflow.com*. Retrieved from http://stackoverflow.com/questions/21737613/image-of-hsv-color-wheel-for-opencv

Steadman, A. C. (2011, December 16). *Applying neuroscience to enhance tactical leader cognitive performance in combat*. U.S. Army Command and General Staff College, Fort Leavenworth, KS. Retrieved from http://www.dtic.mil

Stensrud, B., & Hamel, L. (2015). *The Tactical Execution System (TXS) (Formerly ASIDE)*[PowerPoint]. Retrieved from https://cle.nps.edu/portal/site/7f1e1b96-4c52-4072-a002-95921c66317f/page/6b1cdc2d-403e-4ca2-b7d5-3fed5deb801c

Stout, B. (2000). The basics of A* for path planning. In *Game Programming Gems*, *1*, 254–262.

Straatman, R., van der Sterren, W., & Beij, A. (2005). Killzone's AI: Dynamic procedural combat tactics. In *Game Developers Conference*.

Szalma, J. L., Teo, G., Hancock, P. A., & Murphy, J. S. (2011). Knowledge of results and diagnostic power: Implications for vigilance training to support improvised explosive device detection. In *Proceedings of the Human Factors and Ergonomics Society 55th Annual Meeting*, *55*(1), 1548–1551. doi: 10.1177/1071181311551322

Tozour, P. (2003). Search space representations. *AI Game Programming Wisdom*, *2*(1), 85–102.

Ubink, E., Aldershoff, F., Lotens, W., & Woering, A. (2008). Behavior modeling through CHAOS for simulation of dismounted soldier operations. In D. A. Trevisani (Ed.), *Modeling and Simulation for Military Operations III* (Vols. 1–6965). SPIE Defense and Security Symposium: International Society for Optics and Photonics. Retrieved from http://proceedings.spiedigitallibrary.org/

Ubink, E. M., Lotens, W. A., & Aldershoff, R. F. (2008). *Models & methods in SCOPE: A status report* (Report No. TNO-DV-2008-A126). TNO Defence Security and Safety Soesterberg (Netherlands). Retrieved from http://www.dtic.mil

United States Army Research Laboratory. (n.d.). Improved Performance Research Integration Tool [Webpage]. Retrieved from http://www.arl.army.mil/www/default.cfm?page=445

United States Marine Corps. (2000). *Combat Stress* (Marine Corps Reference Publication 6–11C). Washington, D.C.: United States Marine Corps. Retrieved from http://www.au.af.mil/au/awc/awcgate/usmc/mcrp611c.pdf

United States Marine Corps. (n.d.). Improvised Explosive Device (IED) W3H0005XQ Student Handout. The Basic School, Marine Corps Training Command. Retrieved from http://www.trngcmd.marines.mil/Portals/207/Docs/TBS/W3H0005XQ%20Improvised%20Explosive%20Device.pdf?ver=2016-02-12-073759-177

Unity. (2017). Documentation [Online Documentation]. Retrieved from https://docs.unity3d.com/550/Documentation/ScriptReference/

Unity Technologies. (2016). Unity (Version 5.5.0f3 Personal 64-bit) [Computer Program].

van der Sterren, W. (2001). Terrain reasoning for 3D action games. In *Game Programming Gems* (Vol. 2, pp. 307–323). Retrieved from http://www.gamasutra.com/features/gdcarchive/2001/vandersterrenw.pdf

van der Sterren, W. (2002). Tactical path-finding with A*. *Game Programming Gems*, *3*, 294–306.

Van Nostrand, S. J. (1986). *Model effectiveness as a function of personnel [ME=f(PER)]* (Study No. CAA-SR-86-34). U.S. Army Concepts Analysis Agency. Retrieved from http://www.dtic.mil

Wellbrink, J. C. G. (2003). *Modeling reduced human performance as a complex adaptive system* [Dissertation]. Naval Postgraduate School, Monterey, CA. Retrieved from http://hdl.handle.net/10945/9885

Wichmann, B. A., & Hill, I. D. (1987). Building a random number generator. *Byte*, *12*(3), 127–128.

Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2013). *Engineering psychology and human performance* (Fourth). Pearson.

Wing, V. F. (2002). Avoiding "Ready, Shoot, Aim": An alternate view of beyond the boundaries. *Phalanx*, *35*(4), 26–28.

Wojtowicz, R. (2013). *VBS AFS (Advanced Fatigue System) synopsis and implementation*. Design Department, Bohemia Interactive Simulations.

Woodill, G., Barbier, B., & Fiamingo, C. (2010). *Closing the gaps-Taking into account the effects of heat stress and fatigue modeling in an operational analysis* (Report No. RTO-MP-HFM-202). Retrieved from http://www.dtic.mil

Wray, R. E., Laird, J. E., Nuxoll, A., & Jones, R. M. (2002). Intelligent opponents for virtual reality trainers. Retrieved from http://s3.amazonaws.com/academia.edu.documents/43765263/ITTSEC2002.pdf? AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1493852177&Sig nature=QQ7FRu87OtKi%2BoJNO%2BJ8Hb4R1Ag%3D&response-content-disposition=inline%3B%20filename%3DIntelligent_opponents_for_virtual_realit. pdf

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, *18*, 459–482.

Zacharias, G. L., MacMillan, J., & Van Hemel, S. B. (Eds.). (2008). *Behavioral modeling and simulation: From individuals to societies*. Washington, D.C.: The National Academies Press.

THIS PAGE INTENTIONALLY LEFT BLAK

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
    Ft. Belvoir, Virginia

2. Dudley Knox Library
    Naval Postgraduate School
    Monterey, California